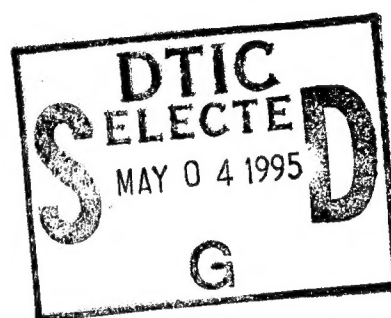AFIT/GOA/ENS/95M-03

DTIC
SELECTED
MAY 0 4 1995
G

APPLICATIONS OF

STATISTICAL PROCESS CONTROL IN

MONITORING AIRCREW BOMBING PROFICIENCY

THESIS

Kirk G. Horton, Captain, USAF

AFIT/GOA/ENS/95M-03

Approved for public release; distribution unlimited

19950503 102

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 1995	Master's Thesis

4. TITLE AND SUBTITLE

APPLICATIONS OF STATISTICAL PROCESS CONTROL
IN MONITORING AIRCREW BOMBING PROFICIENCY

5. FUNDING NUMBERS

6. AUTHOR(S)

Captain Kirk G. Horton, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology, WPAFB OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GOA/ENS/95M-03

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The current tools used by squadron supervisors to monitor the bombing performance of aircrews flying F-111E aircraft are monthly reports that have little predictive capability. No real-time methodology exists for monitoring and predicting aircrew bombing performance and preventing bombing problems that might cause an individual to become unqualified. It has been suggested that Statistical Process Control (SPC) can be applied to the bombing process to develop tools for managing the process, correcting problems, and improve the bombing performance of a squadron.

This study investigates the application of SPC to the bombing process. It examines data taken from an F-111E Fighter Wing during a sixth-month training period. The goal is to develop a control charting scheme that is both useful to squadron supervisors as well as simple to apply by squadron weapons officers.

The results indicate that SPC methodologies can have a significant impact on the bombing process. Control charts generated from the data can give insights to the bombing performance and capabilities of individual aircrews. These insights can lead to improvements in the bombing performance of individuals, as well as in the bombing performance of their squadron as a whole.

14. SUBJECT TERMS

Fighter Aircraft, Bomber Aircraft, Weapons Delivery, Bombing, Statistical Process Control, Control Charts

15. NUMBER OF PAGES
74

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

AFIT/GOA/ENS/95M-03

APPLICATIONS OF STATISTICAL PROCESS CONTROL IN MONITORING

AIRCREW BOMBING PROFICIENCY

THESIS

Presented to the Faculty of the Graduate School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Operations Research

Accesion For		
NTIS CRA&I		X
DTIC TAB		☐
Unannounced		☐
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

Kirk G. Horton, B.E.

Captain, USAF

MARCH, 1995

THESIS APPROVAL

STUDENT: Capt. Kirk G. Horton   CLASS: GOA-95M

THESIS TITLE: APPLICATIONS OF STATISTICAL PROCESS CONTROL
IN MONITORING AIRCREW BOMBING PROFICIENCY
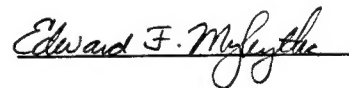
DEFENSE DATE: 7 March 1995
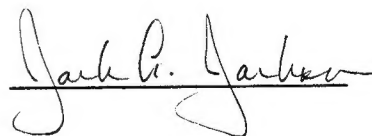
COMMITTEE	NAME/DEPARTMENT	SIGNATURE
ADVISOR:	Edward F. Mykytka, Ph.D. Associate Professor of Operations Research Department of Operational Sciences	*Edward F. Mykytka*
READER:	Jack A. Jackson, Ph.D., Lt Col, USAF Assistant Professor of Operations Research Department of Operational Sciences	*Jack A. Jackson*

Preface

Currently, virtually all flying squadrons that drop practice bombs for aircrew proficiency have only one tool for measuring the state of an individual's bombing process. That tool is a report of past performance by the aircrews. By its nature, the report has limited utility, as it provides little or no short term prediction of performance.

There exist techniques in Statistical Process Control (SPC), a branch of Quality Control, that can provide continuous prediction and indication of the state of the bombing process. These techniques require extensive statistical examinations of the data to determine their suitability.

The purpose of this study then, is to examine representative data from the bombing process of the F-111E to determine if the techniques of SPC can be effective tools for monitoring the bombing process. If the techniques are applicable, the study will describe a strategy for handling and plotting data to maximize the benefits from the SPC techniques.

I am indebted to many persons for their assistance in this research. First, I thank Maj. (Lt Col select) Marshall C. Miller, a former crewmate in F-111s and graduate of AFIT, for giving me the idea for the research and the guidance on how to collect the necessary data. Secondly, I thank my advisor, Dr. Ed Mykytka, who taught me everything I know about SPC and Quality, and who endured my fighter jargon and bomb-talk with a bemused look on his face. Without his instruction and guidance in statistical hypotheses, I would not have gotten very far. Third, thanks to Dr. (Lt Col) Jack Jackson, my reader and fellow F-111 combat alumnus. The words contained herein benefited greatly from his assistance as a bridge between the fighter world and the academic world. Lastly, I would like to thank Dr. Dan Reynolds, who taught our entire class linear models and statistics like no one else could.

Finally, to my wife Susan, we both learned a lot during these 18 months. Most importantly, we learned that we are, and always will be, each other's best friend. I hope the next 18 months are as fun.

Kirk Gerritt Horton

Table of Contents

List of Figures

List of Tables

Abstract

The current tools used by squadron supervisors to monitor the bombing performance of aircrews flying F-111E aircraft are monthly reports that have little predictive capability. No real-time methodology exists for monitoring and predicting aircrew bombing performance and preventing bombing problems that might cause an individual to become unqualified. It has been suggested that Statistical Process Control (SPC) can be applied to the bombing process to develop tools for managing the process, correcting problems, and improve the bombing performance of a squadron.

This study investigates the application of SPC to the bombing process. It examines data taken from an F-111E Fighter Wing during a sixth-month training period. The goal is to develop a control charting scheme that is both useful to squadron supervisors as well as simple to apply by squadron weapons officers.

The results indicate that SPC methodologies can have a significant impact on the bombing process. Control charts generated from the data can give insights to the bombing performance and capabilities of individual aircrews. These insights can lead to improvements in the bombing performance of individuals, as well as in the bombing performance of their squadron as a whole.

APPLICATIONS OF STATISTICAL PROCESS CONTROL IN MONITORING AIRCREW BOMBING PROFICIENCY

1. Introduction

1-1. Aircrew Bombing Proficiency

Aircrew Bombing Proficiency (ABP) is a pilot's or weapons system operator's (WSO's) ability to hit a specified target with a projectile released from his/her aircraft, after he/she has learned the basic skills required to perform the task. For the vast majority of aircrews (pilots/WSOs) in the fighter/bomber force, maintaining bombing proficiency requires frequent and repeated practice. In the USAF, sorties dedicated to maintaining or improving an aircrew's proficiency in various events are called continuation training (CT) sorties. Aircrews typically get between four and twenty CT sorties each month, depending on aircraft type. For those aircraft whose primary wartime mission is bombing, the focus on most CT sorties is maintaining ABP.

Bombing events are distinguished by two delivery parameters: dive or climb angle and system dependence (automatic/manual release). Thus, diving/manual deliveries and climbing/automatic deliveries, among others, are counted as distinct bombing events. Aircrews flying CT sorties receive a score for each bomb released, which indicates both the distance and direction from the target.

Fighter wings use bomb data in several basic ways, depending on the particular

wing's policy. Typical usage includes:

- End-of-month Top Gun (best bomber) awards (based on average miss distance for various bombing events during the month);

- End-of-month bombing proficiency reports (broken down by average miss distance for each event for the month);

- End-of-quarter Top Gun awards (same as above, based on calendar quarter);

- End-of-half Top Gun awards (same as above, based on calendar half);

- End-of-half bombing qualification.

For bombing qualification, an aircrew must hit at least fifty percent of his/her targets in mandatory bombing events. Hit criteria, or maximum allowable miss distances (also called hat sizes), are based on aircraft type and delivery parameters. Mandatory events are also determined by aircraft type and include all deliveries which the specific fighter wing would be required to perform in a wartime scenario. As implied by the preceding list, bombing qualification is a semi-annual requirement, monitored monthly through proficiency reports. Should an aircrew begin to show a lack of proficiency, as captured by a monthly report, supervision and instruction are increased to ensure end-of-half qualification. Should an aircrew fail end-of-half qualification, he/she must undergo requalification training. Requalification training requires more sorties to be flown, wasting resources that would otherwise be spent on CT, and is therefore an undesirable occurrence; however, nothing is done at the local level to continuously predict when an aircrew will fail to qualify in bombing. While monthly reports can capture a lack of proficiency, they cannot predict the continuation of that condition, nor show any trend that may be present in an aircrew's bombing performance.

Typically, new tools for managing a process are developed from studies of the process undertaken at higher-headquarters levels. Studies of the bombing process would normally be performed by either a Pentagon or an Air Combat Command (ACC) analysis

team, or a contractor working for one of those teams. A search of the Department of Defense (DOD) databases should reveal previous studies of the bombing process, and whatever new tools were developed from those studies. However, only one study related specifically to ABP can be found. Undertaken by the Air Force Center of Study and Analyses (the Pentagon) in 1984-86, the analysts used a mathematical model to relate ABP to flight hours per month, sorties per month, and bombing events per training cycle. The model developed has several purposes, but it is not able to predict when an individual aircrew would no longer be proficient enough in bombing to qualify at the end of a half.

By expanding the search criteria to include all munitions delivery proficiency training, an additional study was discovered. Sponsored in part by the Naval Postgraduate School (NPS), the study applied Statistical Process Control (SPC) techniques to Naval Gunfire Support (NGFS) training. The authors showed SPC techniques to be valuable tools for monitoring and improving the NGFS process. While no application of SPC methodologies to ABP specifically have been found, the many parallels that exist between ABP and NGFS hold promise for the application of SPC to ABP.

Currently, in spite of massive amounts of bombing data, no tools have been developed to predict when aircrews will no longer be proficient in bombing. The central question is: Does prediction of such an event serve a useful purpose? The answer, of course, will determine the value of this research. The author contends that *prediction* of the lack of ABP enables the *prevention* of that condition. A device that continuously tracks ABP and signals when an aircrew is experiencing difficulties or when an upward trend in his/her scores exists, would be invaluable in the prevention of qualification failures and the detection of poor bombing habit patterns. An aircrew experiencing such problems could be supervised more closely and given more instruction *before* he fails to

qualify for the half and before he develops poor habit patterns on the range. This provides immediate benefits for each aircrew, as well as long run benefits for the Air Force. These benefits take the form of more consistent bombing overall and less requalification training, leaving more CT sorties and dollars for other urgent training requirements.

An example should serve to illustrate this concept. In January of 1992, a WSO with whom the author flew infrequently, began to develop a problem in his aimpoint on radar targets. He was using the radar incorrectly when updating the navigation system in such a way that most of his bombs fell short. After an extensive period of short bombs he began to compensate by aiming long. His problem wasn't discovered until a foggy morning in April when all local takeoffs were delayed. On such weather days, aircrews normally gathered in a briefing room and talked about techniques in various flight events. On that day the discussion was about aiming on the range, and that WSO had a revelation. Luckily, it was early enough in the half to save his qualification (the monthly reports failed to catch his problem due to his aiming compensation), but if his problem had developed any later in the half salvage would have been impossible. His radar error may have cost him his bombing qualification before he could develop a compensating habit, and the squadron would have been forced to spend scarce training dollars to requalify him in bombing.

Systematic bad habit patterns are quite common among fighter/bomber aircrews. Unfortunately, compensations are just as common, so that very few problems are discovered and corrected before doing substantial damage to training budgets. Had a bombing proficiency tracking system been in place in the above situation, the WSO may not have had a chance to develop a bad habit. His problem would have been discovered much sooner, and corrective/preventive action could have been implemented to reinforce a good habit pattern on the range. After all, to discover and remove an assignable cause

for poor bombing is better than to compensate for it. Clearly, the Air Force needs a better way to treat bombing data to continuously predict ABP and prevent bombing problems from causing an aircrew to become unqualified for a flying half.

1-2. Approach

With the goal of developing a tool for continuously monitoring the bombing process, and the precedent from the NGFS study, the author will apply Statistical Process Control (SPC) techniques to an actual F-111E bomb score database. The analysis will use several types of control charts to investigate the application of SPC in ensuring ABP. The overall goal will be to determine whether SPC techniques can be used to continuously predict ABP and prevent bombing problems that cause aircrews to become unqualified.

While many bombing events exist, only two will be examined. The events to be included in the analysis are:

Visual Laydown Delivery (VLD) -- pilot-controlled level delivery;

Radar/System Laydown Delivery (RLD/SLD) -- WSO-controlled level delivery.

The first part of the analysis will be focused on determining the most effective way to chart the data. Different types of data require specific control charts to correctly display the properties of the process. For example, highly correlated data require modeling to remove the correlation before charting, while data from different populations should not be plotted on the same chart. Only by careful analysis of the actual data can decisions about how to chart the data be made. Various statistical tests and visual depictions of the data will be used to justify how the data is grouped for control charting. Once the data

have indicated the most effective charting strategy, the choice of control charts will be made.

After determining the control charts to use, the analysis will turn to applying those charts to the data. The author will set up control charts for each aircrew in the database, compute control limits based on specific data point exclusion rules, and chart the data. The chi-square goodness-of-fit test will be used to estimate the underlying distribution in the data. That distribution will be used to estimate the properties of the control charts; these properties indicate the effectiveness of the charts in monitoring the process.

The last part of the analysis will focus on evaluating the approach based on the aforementioned overall goal: how effective are the tools developed using SPC techniques in continuously predicting ABP and preventing bombing problems from causing an aircrew to become unqualified? To answer this question, the author will evaluate the usefulness of the approach, the impact of the control limits for each chart, and whether SPC techniques could have prevented problems encountered by aircrews during the period represented by the data.

1-3. Scope

This analysis examines only F-111E bomb score data from a single wing (20th Fighter Wing, RAF Upper Heyford, U.K.) accumulated over a six month period. The data includes complete information (for six months) on approximately 180 pilots/WSOs who released between ten and thirty bombs in each event (three events each). While this might imply limited application of the results, the methods to be used can be applied to any squadron that drops practice bombs and/or *tones* (simulated bombs). Extensions can

also be made to units that perform scored target practice. In fact, SPC can be an effective tool for any process in which individual performance is similarly scored.

1-4. Assumptions

Studies involving actual data are notoriously more difficult than those using simulated information; data may not be representative of the population, and worse, may not come from a single underlying distribution. Since this study uses actual fighter wing data, assumptions must be made regarding these characteristics. First and foremost, it must be assumed that out of the data set, samples can be obtained that are representative of the distribution of ABP across the 20th Fighter Wing. Since the data consists of the *entire population* of information on ABP for the wing for six months, this assumption of representativeness is not unreasonable.

Another significant assumption that must be addressed is one of homogeneity. To justify plotting the scores attained by aircrews flying different aircraft over different bombing ranges together on a single control chart, the distributions within each category (aircraft or range) should be as homogeneous as possible. If they are not, variations between categories could mask variations within each category, caused by an out-of-control condition. In this case, the resulting control limits would be too wide to effectively monitor the process. In this analysis, extensive statistical tests are employed to demonstrate the homogeneity of the data. If such SPC techniques are used in the field, the assumption of homogeneity may have to replace the tests used herein if the user does not have access to sufficient data or similar statistical tests. This assumption is reasonable as long as their exist no obvious reasons to believe individual aircraft or ranges differ from each other.

1-5. Results

The results of this research have a twofold application: direct application to F-111 fighter squadrons, and indirect application to all bomb dropping squadrons, as well as to other units with similar processes. Appendix B is a "How to" guide for use in F-111 squadron weapons shops. It describes how to implement the control charting procedures outlined in subsequent chapters. Once published, the guide can also be adapted for use in bomb dropping squadrons of other aircraft types. With these two applications, the results of this research have the potential for far reaching effect.

2. Previous Munitions Delivery Proficiency Studies

Munitions delivery training programs have enjoyed a long history. One would suppose that studies of the data from these programs would abound. Quite surprisingly however, an extensive search for studies has produced very few. It would appear that gunnery and bombing training methods have evolved due mostly to trial and error or tradition rather than on the basis of analytical results. Even so, at least one study using a traditional statistical approach to aircrew bombing proficiency exists, undertaken by the Air Force Studies and Analysis Agency (AFSAA). The first part of this chapter reviews pertinent areas of this study.

While the *quality movement* has been in full swing in Japan since shortly after the end of World War II, it has taken quite a bit longer to take hold on this side of the Pacific. The comparatively slow rate of acceptance of quality improvement methods in the United States has limited the number of previous analyses of munitions delivery training and proficiency using SPC methods. One recent study of Naval gunnery proficiency is of particular interest. The second part of this chapter extracts methodology and results from that study and applies them to the current analysis.

2-1. The Proficiency Issue

An "Analysis of Factors Affecting Pilot Proficiency" was conducted in 1984 by the Fighter Division, Directorate of Theater Force Analyses, Air Force Center of Studies

and Analyses, now called Air Force Studies and Analysis Agency (AFSAA). The purpose of the study was to model aircrew bombing proficiency as it relates to flying frequency and experience. Data for the analysis consisted of bomb score records covering one year from two A-10 and two F-16 squadrons (2:1). While direct application of the results to the current analysis are limited due to the scope of the study (i.e., no F-111E information), the results give some general information about possible sources of variation in data. The following sections summarize the key results of the study. Each summary is accompanied by a description of its relevance to the current study.

The Experience Factor. The pilot proficiency study found that mission flying experience (i.e., time in fighters) has the highest correlation (of all factors considered) with bombing accuracy; the more experience an aircrew has, the better he bombs. The report notes that although experience in the aircraft is correlated with bomb scores, flying frequency is not, despite the fact that the accumulation of experience is dependent on how often an aircrew flies (2:9).

The correlation between experience and bombing accuracy impacts the current study in an indirect way. As will become clear in the next chapter, the success of SPC techniques in ABP relies on proper subgrouping of data. Statistical tests performed on the F-111E data will show that all aircrews in the study can be grouped together when computing control chart parameters. However, because experience information is not available from the data, these tests ignore the experience levels of the aircrews in the database. If differences exist between inexperienced and experienced aircrews, but the two groups are taken together, control limits can become distorted and less useful. While this fact can affect the control limits computed herein, practically speaking, experience levels will be known by the squadron applying SPC techniques, which will allow the

squadron to test for homogeneity between the two groups before computing control limits.

While data homogeneity is important, the implication of the lack of correlation between flying frequency and bombing proficiency is also significant. This fact, if translatable to the F-111E, means that the same control charts and control limits can be applied to all aircrews regardless of how frequently they fly. Universally-applicable control charts would greatly reduce the workload of process monitors, and promote the use of an SPC package in the squadron.

Predicting Proficiency. The authors of the pilot proficiency study developed a model that relates pilot capability to mission experience. The model shows that certain experience thresholds exist above which a marked increase in capability occurs (900 hours for F-16, 1400 hours for A-10). The model can also be used to predict proficiency at given flying frequencies (the rate at which a pilot is accumulating mission experience.) (2:13)

If these results could be translated to the F-111E and a similar model developed to express the relationships, the experience threshold obtained can be used to differentiate between inexperienced and experienced aircrews when computing control chart parameters. The relationship between flying frequency and bombing proficiency could then be used in conjunction with the control charts to develop training programs. Such training programs would be the beginning of the bombing improvement process for a squadron.

2-2. Statistical Process Control Applications

In 1993, a study of the management of Naval Gunfire Support (NGFS) using SPC methodologies was undertaken. NGFS is the process whereby a ship acquires a surface target, loads its guns, then aims and fires at the target while navigating and possibly executing wartime tactics. The goals of the study were threefold: to describe how SPC techniques can apply to the NGFS process, to highlight practical hurdles in applying the methodology, and to discuss the benefits to NGFS of the application of SPC techniques (1:5). The parallels which exist between this NGFS study and the present ABP analysis are considerable. The next sections describe results from the NGFS study and how they relate to the current study.

SPC Methodology. The primary contribution provided by the Navy paper is precedence for the use of short production run SPC techniques. The authors collected data from four sets of engagement exercises, consisting of 72, 82, 48, and 115 observed firings. From each observation set, only a small fraction were fired for a score; thus only that fraction from each set was usable test data. Other firings were for acquisition, spotting, and boresighting. The authors applied a moving average chart and a moving range chart (collectively known as mA/mR charts) to data collected from the NGFS process. The extremely limited nature of their usable database justified the use of the mA/mR charts. (1:6)

The Naval gunnery process is very similar to aircrew bombing training. Aircrews often drop only one or two bombs on a range for each event, limiting the amount of data in each sample and precluding the use of standard control charts which require larger samples to be effective. Discussion of the specific charts that can most effectively present the data in the current study is deferred to Chapter Three.

Subgrouping.   The second contribution from the Navy paper is a discussion of subgrouping. At issue is whether the authors could assume all ships' scores were from the same population in computing control limits. The authors use Q-Q plots (3:48) and K-S bounds (4:140) to conclude that aggregation of the data is appropriate; however, they also demonstrate how such aggregation lowers control chart effectiveness by widening control limits. Their standard Xbar chart obtained by aggregating the data has very wide control limits, and is unable to identify an out of control condition previously identified by an mA/mR chart combination (1:10). The combination of statistical tests and shortcomings of standard control charts support the use of short production run techniques and individually computed control limits for this scenario.

The parallel between NGFS and ABP is quite clear in the issue of subgrouping. As will be seen with the data for the current study, many areas for possible non-homogeneity exist in aircrew bombing training. Each aircrew flies a different aircraft to a different range for each sortie. These aircrew, aircraft, and range categories could each conceivably possess a unique distribution, and aggregating data from them would be inappropriate. The next chapter investigates each of the possibilities in the search for the right control charts and limits for the ABP process.

Other Applicable Issues.   The Navy researchers had the goal of matching NGFS war plan requirements (i.e., specifications) with ships' process capabilities (1:10). This goal is in line with one of the goals of the current study, to get a better gauge of aircrew bombing proficiency than event hat size.

In the Navy research effort, the researchers were the data collectors, and so could document assignable causes as they occurred and exclude data points as necessary (1:7). They also molded the data collection process into exactly what they needed. The current study allows neither opportunity. This author must accept the data as it is and do without

unavailable information. It will be shown that this shortfall can seriously affect decisions made about the inclusion of data points and the calculation of control limits.

2-3. Summary of Previous Efforts

Studies using traditional methodologies appear to be few and far between, and when found, quite tangential to the SPC methodology proposed. Even so, results can be useful to the current study when it is noted that different distributions in a population may be present. This heads up to possible problems is the contribution made by the Pilot Proficiency study produced by AFSAA.

SPC methodologies, as applied to defense issues, are still in their infancy:

"COMNAVSURFLANT [Commander, Naval Surface Fleet, Atlantic] and AFWTF [Atlantic Fleet Weapons Training Facility] agreed that our novel application of SPC methodology was promising...the QMB [Quality Management Board] also agreed that studies of SPC should be pursued in other defensive warfare applications." (1:11)

Clearly, the Navy study is perceived as a unique perspective. Precedents gleaned from this study, such as useful SPC techniques and supportive statistical tests, can be invaluable to a follow on study such as this one.

3. Preliminary Analysis of Data and Calculation of Control Limits

Having established the possible application of SPC techniques to munitions delivery training programs, it is now time to examine the data and make several preliminary decisions about control charts. As mentioned in Chapter One, the data used herein is six month's worth of bomb scores by F-111E aircraft and aircrews from the 20th Fighter Wing during 1992. The data contain information on at least five different bombing events. This study focuses on two of those events, visual level bombing (VLB) and radar laydown delivery (RLD). These events are the bread and butter of the F-111 mission, and are similar to each other. Both require an extremely low, fast approach to the target. VLBs are delivered using outside references for bomb release, while RLDs are delivered using radar cues and bombing/navigation system range calculations for bomb release.

The data also contain information on twelve different bombing ranges and over twenty different individual aircraft. Scores in each event were recorded for over thirty individual aircrews. Two central issues to the effectiveness of control charts of the data are how to chart all of this data and how to estimate the process means and standard deviations needed to compute control limits. First, what is the most effective way to chart the data? Should all aircraft be on the same chart, or all ranges? Since the intent is to monitor individual aircrew's bombing proficiency, it is assumed each aircrew should be charted separately; but are there differences between the scores attained with individual aircraft or on certain ranges such that they have to be charted separately as well? Second, what is the best approach to estimating the process means and standard deviations? Which points in the data can be considered representative of the underlying

distribution? Which points can be determined to be outliers, and should those outliers be excluded from control limit calculations? These are the questions addressed by the following analysis.

This chapter and the next trace the preliminary data analysis required to determine which types of charts represent the data most effectively, and how to estimate the process means and standard deviations. The tools used in the analysis are the paired-t test, Bartlett's correlation test, and the analysis of variance (ANOVA). Relative frequency histograms and Q-Q plots are used to visually support results from the tests. Once the data have indicated the best charting strategy, process means and standard deviations are estimated, and from these estimates, control limits are calculated. Several approaches to excluding points from control limit calculations are discussed, and the most effective control limits are chosen to be used on the control charts in Chapter Four. Chapter Four begins with theoretical distribution fitting of the data to predict probabilities associated with the charts (e.g., probability of a point plotting out of control when the process is actually in control: a false alarm). These probabilities are then compared to empirical results from the data to determine the effectiveness of the control charts. Finally, the data are plotted on control charts and examined for tell-tale signs of process performance.

3-1. Preliminary Data Analysis

This section describes each statistical test applied to the data, the reason for using it, and the results obtained from the data for each bombing event. First are the tests for data correlation, the presence of which could impact the selection of control charts. Following these tests are the ANOVAs used to determine if separate charts are required for each range and/or tail number. The last tests in the section are the ANOVAs required

to determine if control limits can be calculated for all aircrews, or if individual control limits are required for each aircrew.

Choosing Control Charts. Since most control charts are designed for constant processes, the bombing process must be assumed to be constant. As the results of the AFSAA study show, bombing proficiency increases with the build-up of mission experience, a process so slow as to be essentially constant for the purposes of this study. Later, the lack of short term (six months) trends in the bombing process will be shown.

The data is for an entire squadron of aircrews. It contains an average of 25 VLB scores each for 32 pilots, and an average of 21 RLD scores each for 32 WSOs. The scores are arranged chronologically by aircrew, range, and tail number. Aircrews drop one or two bombs on the same target during a sortie. Because the number of bombs dropped during a sortie is not constant and is so small, control charts which rely on larger, constant sample sizes are inappropriate. In situations where the logical sample size is one and no correlation exists between samples, Wheelers & Chambers recommend the X chart for individual measurements paired with a moving range (mR) chart with a span of two (7:217). This XmR pairing is the set of control charts to be initially used in this study. Since XmR charts and their control limits assume independence and identical distribution of the data, tests for the equality of means and the lack of correlation between successive passes will be performed in the next section in order to justify the use of these charts. It is assumed that this type of correlation is the only type that may be significant in the data.

Paired-t Test for First and Second Pass Means. In order to plot both first and second pass scores on the same control chart, the scores must at least come from distributions with the same mean. Practically speaking, one would expect that the score for the second pass on a target to be lower than the first pass, because of corrections applied by the aircrew. If the difference in means can be shown to be small or non-

existent, all bomb scores from the same event can be assumed to come from identical distributions (since there is no reason to believe that the distributions take on different functional forms.)

To show equality of means, the paired-t test is applied to those scores which are part of a two-pass sequence. The t test is applied to the signed differences between each pair. If the test statistic is not significant, then the means can be concluded to be equal.

VLB Results. After omitting unpaired data points, the data was arranged into 187 pairs. As shown in Figure 3.1, it can be concluded that these pairs have equal means at a significance level of 0.107. While this is somewhat surprising, it is nonetheless fortunate, since straightforward SPC techniques can still potentially be used on the data.

Figure 3.1 Paired-t Test Results (VLB Scores)

	fra	second	t Stat	P(T<=t)	t Crit
Mean	129.1872	114.7005	1.246406	0.107091	1.653088
Observations	187	187			
Pearson Correlation (r)		0.304459			
Hypothesized Mean Difference		0			
df		186			

RLD Results. After identical setup of the data, 161 pairs were obtained. Figure 3.2 shows that it can be concluded that these pairs also have equal means at a significance level of 0.264.

Figure 3.2 Paired-t Test Results (RLD Scores)

	fra	second	t Stat	P(T<=t)	t Crit
Mean	124.0062	133.2236	-0.632721	0.263909	1.654432
Observations	161	161			
Pearson Correlation (r)		0.323992			
Hypothesized Mean Difference		0			
df		160			

Bartlett's Test for Correlation. Even if first and second pass scores come from identical distributions, they may still be correlated. Highly correlated data tend to shrink control limits and greatly increase the frequency of false alarms on control charts (6:343). A simple test for correlation uses the Pearson correlation coefficient and an estimate of its standard error to compute an acceptance region. If the correlation coefficient is within that region, then the data set can be concluded to be uncorrelated at the significance level of the region (5:368). Figure 3.3 shows the results of the test for both VLB and RLD scores. Since neither correlation coefficient is within the acceptance region, some correlation exists between first and second pass scores; however, the magnitude of the correlation is very low. In order to keep the methodology simple, which will increase the likelihood of it being used by a squadron, this small correlation will be ignored for now. Chapter Four examines the impact of the correlation on the computed control limits and the subsequent performance of the control charts.

Figure 3.3 Bartlett's Test for Significant Correlation

Data Set	Correlation Coeff (r)	95% Acceptance Region
VLB	0.304459	$-0.143329 \leq r \leq 0.143329$
RLD	0.323992	$-0.15447 \leq r \leq 0.15447$

ANOVA. The purpose of control charts is to highlight the variation between samples from a process. Control limits are calculated from the variation within samples; thus, for control limits to be effective, the data must be charted such that variation within a sample is smaller than the variation between samples. By charting only samples from a single distribution on the same chart, variation within each sample is minimized. Control limits calculated based on variation within the samples are thus able to detect abnormally large fluctuations between samples. While X charts for individuals

use a sample size of one, which has no within sample variation, how the data are charted is still very important. The reason is that control limits for X charts are calculated based on the difference between successive samples, variation in which is assumed to be smaller than the difference between samples further apart (7:48).

In the data of this study, there exist many possible causes of variation. It would be impossible to capture all of them, thus many are labeled as *inherent* to the process and are assumed constant across the data. Examples of these causes of variation are: crewing changes (the occupant of the other seat of the jet on a given day), human factors (amount of sleep, diet), and weather (range visibility).

Several other potential causes of variation in the data are more obvious and traceable. These causes can also be identified as those having the greatest impact on charting strategy. The following sections focus on ANOVAs for range to range variation, jet to jet variation, and aircrew to aircrew variation, respectively. The goal of these tests is to determine if scores from all ranges, and/or aircraft can be charted together, and if all aircrew scores can be used collectively to compute control limits. Throughout these tests, it is assumed that distributions take on the same functional forms. It is also assumed that the impact of differences in the means of the distributions far outweigh differences in the standard deviations of those distributions.

Between Range Variability. The data come from twelve different bombing ranges located in various parts of Europe. Each range is operated independently of the others, with different scoring systems and different operators. Because of these facts, the possibility exists that the scores from each range in the data may have come from different distributions. If so, each distribution identified should be considered separately when computing control limits, and each range should be charted on separate control charts. If the analysis does not identify more than one distribution in the data, all

ranges can be grouped together when computing control limits and plotting control charts.

     VLB Results. With VLB scores grouped by range, an ANOVA was run on the data. As Figure 3.4 shows, there is indeed more than one distribution present in the VLB scores. By examining the means in the ANOVA table, it appears that ROSY has a significantly lower mean than all other ranges. Figure 3.5 is another ANOVA table, with all ROSY scores excluded from the analysis. This time the scores appear to come from distributions with the same mean, with a significance level of 0.461. It appears that ROSY range produces lower scores on the average than the other ranges.

Figure 3.4 ANOVA: VLB Scores (by Range)

Groups	Count	Sum	Average	Variance
COWDEN	49	6817	139.1224	19436.86
DONNA	64	8032	125.5	13765.33
HOLB	65	9299	143.0615	25102.25
JURBY	79	9734	123.2152	12084.12
LILST	11	1495	135.9091	5910.491
NORD	5	803	160.6	8663.8
PEMBR	10	1254	125.4	6722.044
SUIPP	3	450	150	2500
TAIN	145	16709	115.2345	13548.63
VLIE	17	1735	102.0588	21206.43
WAIN	135	21591	159.9333	29861.79
ROSY	82	6803	82.96341	8375.962

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	369352.7	11	33577.52	1.910142	0.035257	1.803304
Within Groups	11478787	653	17578.54			
Total	11848140	664				

Figure 3.5 ANOVA: VLB Scores (by Range, without ROSY)

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	184648.1	10	18464.81	0.977921	0.461424	1.847248
Within Groups	10800334	572	18881.7			
Total	10984982	582				

What are the reasons for the lower scores? Lower resolution scoring systems may preclude the operator at ROSY range from distinguishing a 30 foot bomb from a direct hit. The range tower may have inadequate triangulation technology, or may be further from the target than at other ranges. The tower controllers may have a better relationship with the aircrews, and thus may call a lower score than was actually recorded by the measuring system. While there could exist a multitude of such reasons for the lower scores from ROSY, the key is that an assignable cause probably exists which supports the exclusion of ROSY bomb scores from the calculation of control limits. This conclusion is supported by anecdotal evidence gathered through experience at the ROSY range.

While the ANOVA is quite conclusive, a visual depiction of the results is also beneficial. Figure 3.6a is a Quantile-Quantile (Q-Q) plot of ROSY scores vs. all other range scores. Q-Q plots are similar to probability plots except that, instead of plotting the probability (according to the theoretical distribution) on one axis, the plot is created using quantiles from two empirical distributions (3:48). If both quantiles are from the same distribution, the plotted points will approximate a line with a slope of one. The more the plot deviates from this line, the more evidence that one distribution is different from the other, either in mean or in standard deviation.

From the plot, it is clear that the ROSY distribution is different from that for all other ranges. For comparison purposes, Figure 3.6b is the same plot for COWDEN vs. all other ranges. This plot shows the quantiles plotting very near the reference line, indicating that COWDEN scores come from the same distribution as the other ranges. Plots for each of the remaining ranges vs. all others produce similar results (not shown). This evidence further supports the assumption that differences in means are more important than differences in standard deviations for this data, since the Q-Q plots indicate that the distributions have similar standard deviations, as well.

Figure 3.6a Q-Q Plot: ROSY vs. All Others (VLB Scores)
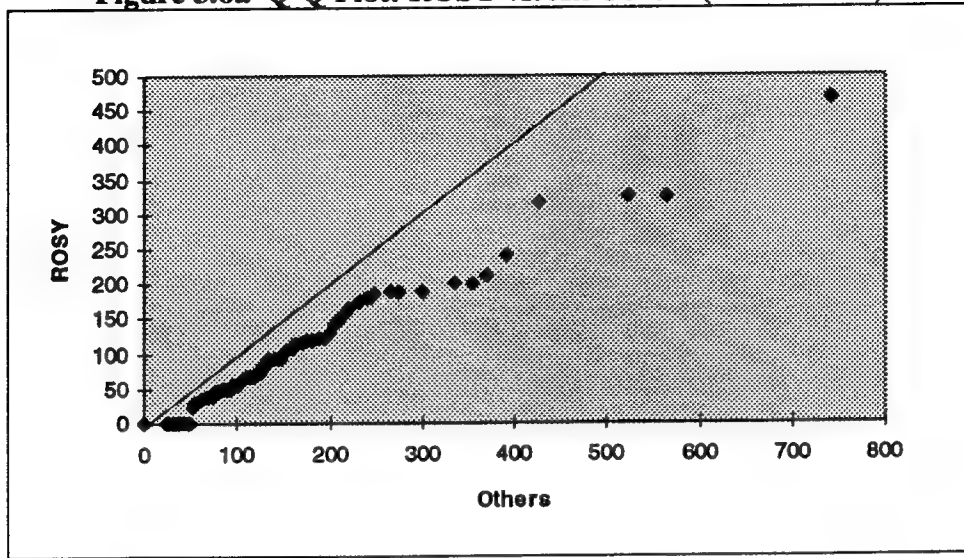
Figure 3.6b Q-Q Plot: COWDEN vs. All Others (VLB Scores)
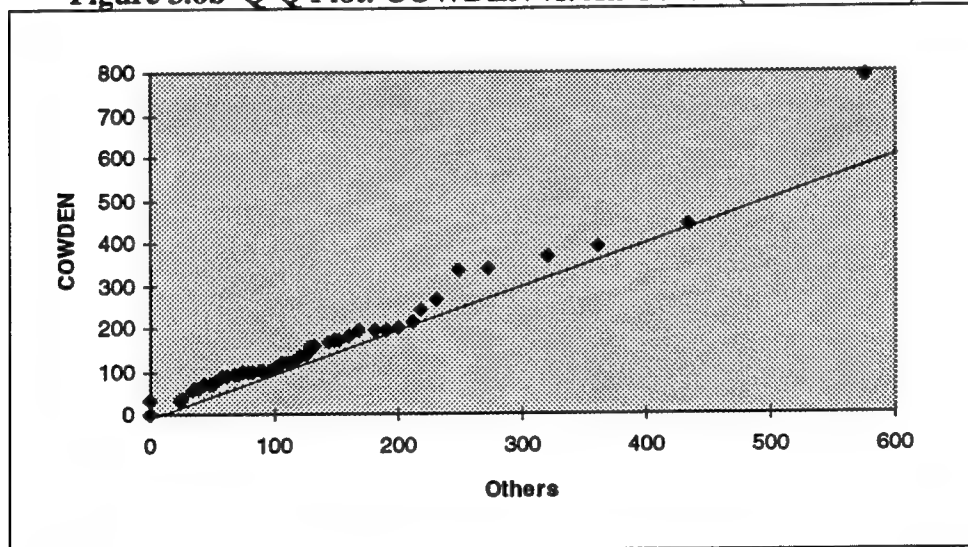
Further visual support is provided by comparing histograms of relative frequencies for ROSY and all other ranges. Figure 3.7 shows this comparison. ROSY appears to have a large spike at zero and a tight distribution around a mean of approximately 75. The other distribution has a smaller spike and a looser distribution around a mean of approximately 135. Clearly, ROSY scores are different and should be excluded from

control limit calculations to avoid lowering control chart effectiveness. While they should also be plotted on separate charts, the extra effort to do so is not worth the marginal improvement in results. ROSY scores will be plotted on the same charts as other ranges, and are not expected to impact the effectiveness of the charts.

Figure 3.7 Relative Frequency Histograms: ROSY vs. All Others (VLB Scores)
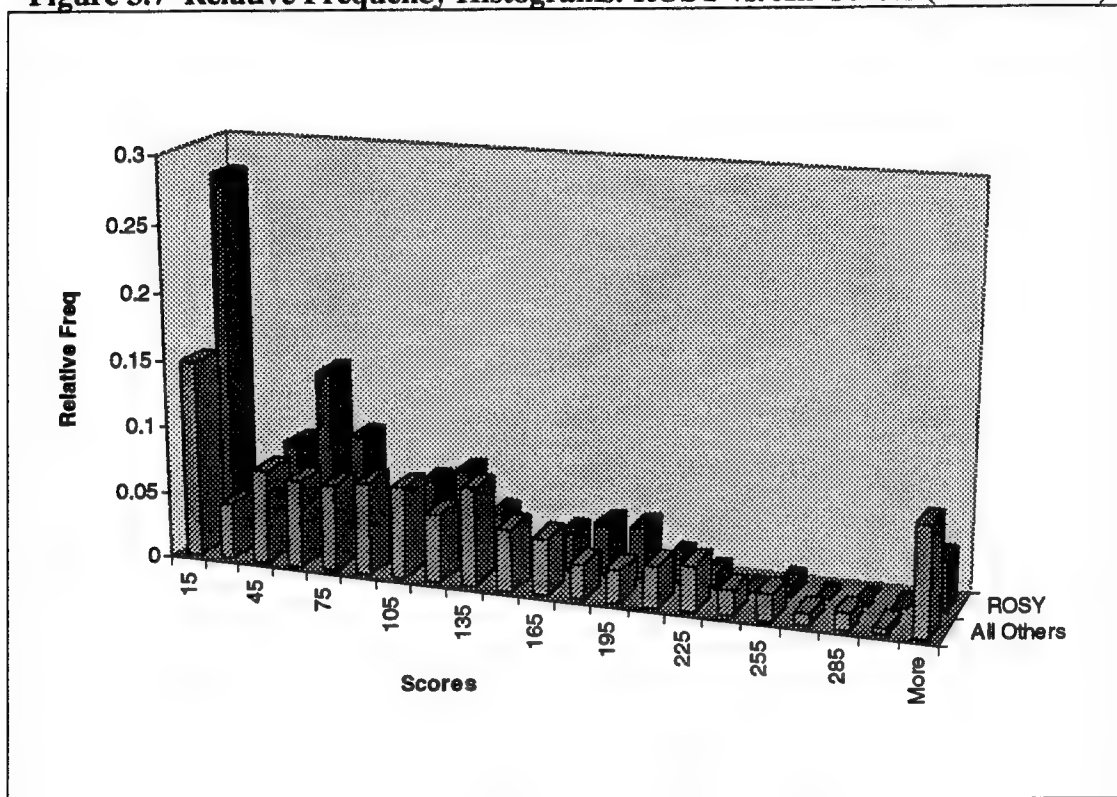
RLD Results. Using the identical procedure, an ANOVA was run on RLD scores grouped by range. Unlike the VLB results, the ANOVA for RLD scores indicates no difference between the means of scores obtained from each range. Figure 3.8 is the ANOVA table for this test.

3-10

Figure 3.8 ANOVA: RLD Scores (by Range)

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	499610.7	12	41634.23	1.516202	0.113985	1.77052
Within Groups	14498641	528	27459.55			
Total	14998252	540				

Between Jet Variability. Variability due to the specific aircraft flown may impact the control limits as well. The data contain scores obtained from 49 different aircraft. Each aircraft is maintained by a different team of crew chiefs, and each has its own unique cycle of system failures and maintenance actions to repair those failures. Because of these possible causes of variation between aircraft, the scores obtained from each one may not come from the same distribution, and may require individually tailored control charts. An ANOVA of the scores in the data grouped by tail number will distinguish these different distributions, if they exist.

Among the 49 aircraft flown by the aircrews in the database, several were flown so infrequently as to make estimations of their distributions ill-advised. For this reason, only jets that dropped more than 20 bombs, either VLB or RLD, are included in the ANOVA. The justification for excluding less frequently flown jets is not for the actual cut-off point of 20 bombs, but more for the reason behind excluding any jets at all. During the time the data was collected, the squadron owned between 18 and 21 aircraft. These jets were most familiar to the aircrews, so the crews knew how to compensate on the range for any of the jets' idiosyncrasies. Contrary to this, jets flown infrequently are borrowed from other squadrons for sorties. Aircrews know very little about these jets, and cannot compensate for them on the range. Due to all the unknown factors involved with these jets, it is advisable to exclude their infrequently sampled data from the ANOVA and control limit calculations. They will, however, be plotted on the same charts (if the ANOVA indicates that all jets can be plotted together.) Out-of-control points produced by non-squadron jets provide insight into possible process improvements (i.e., if more information can be

obtained about these jets by the squadron, maybe fewer out-of-control bombs will be dropped from them.)

VLB Results. By grouping the VLB data by tail number and excluding those jets most likely not owned by the squadron, as well as all ROSY scores, the ANOVA table in Figure 3.9 is obtained. As the figure indicates, no difference exists between jets at a significance level of 0.641. This result is expected since visual bombing is much more dependent on pilot ability than it is on aircraft systems.

Figure 3.9 ANOVA: VLB Scores (by jet)

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	301588.7	20	15079.43	0.858749	0.640791	1.589274
Within Groups	9868594	562	17559.78			
Total	10170182	582				

RLD Results. The same procedure performed on RLD scores yields quite different results. Figure 3.10, the ANOVA table for RLDs, indicates a significant difference exists between the means of scores obtained from different aircraft, with a significance level near zero. The reason is quite obvious and expected. During the period the data was collected, the squadron owned two different models of the F-111E. Analog jets are older and rely on a less than state-of-the-art navigation and bombing system. The newer jets were modified by an avionics modernization program (AMP), and use a superior Global Positioning System (GPS) and ring laser gyro navigation and bombing system. These digital, or AMP, jets have more accurate radar and bombing systems with automatic error compensations for altitude and airspeed. AMP jets should produce better bomb scores by design, and to find otherwise in the data would have been quite a surprise.

Figure 3.10 ANOVA: RLD Scores (by jet)

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1766865	19	92992.9	3.661695	3.67E-07	1.606537
Within Groups	13231387	521	25396.14			
Total	14998252	540				

An ANOVA serves to quantify the difference between analog and digital jets. By grouping aircraft by type, an ANOVA will display the difference between the means of scores obtained in each type. Figure 3.11 is the ANOVA table for this grouping. With a significance level approaching zero, the difference between analog and digital aircraft is clear.

Figure 3.11 ANOVA: RLD Scores (AMP jets vs. Analog jets)

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1158308	1	1158308	45.11059	4.74E-11	3.858759
Within Groups	13839944	539	25677.08			
Total	14998252	540				

Pictorial evidence of the difference in the distributions of the two sets of scores is given in the Q-Q plot of Figure 3.12. As described previously, the further the plot of the quantiles deviates from the line with slope of one, the more one distribution differs from the other. The relative histogram comparison of Figure 3.13 further supports the ANOVA. Clearly, AMP scores are more frequently near zero, with a mean near 110. Analog scores, by comparison, have scores larger than 300 much more frequently, resulting in a mean closer to 200.

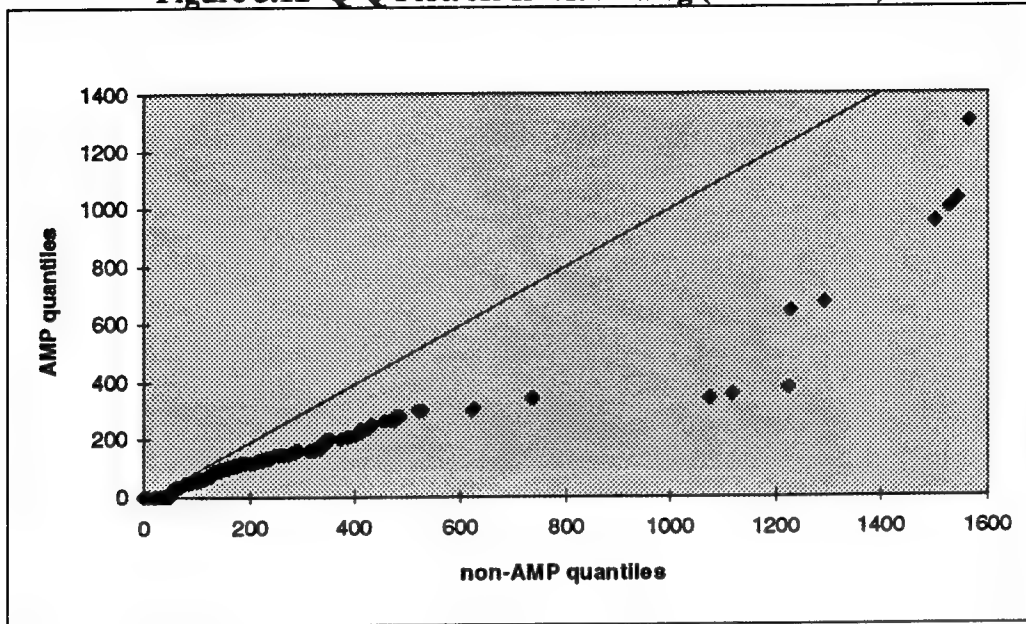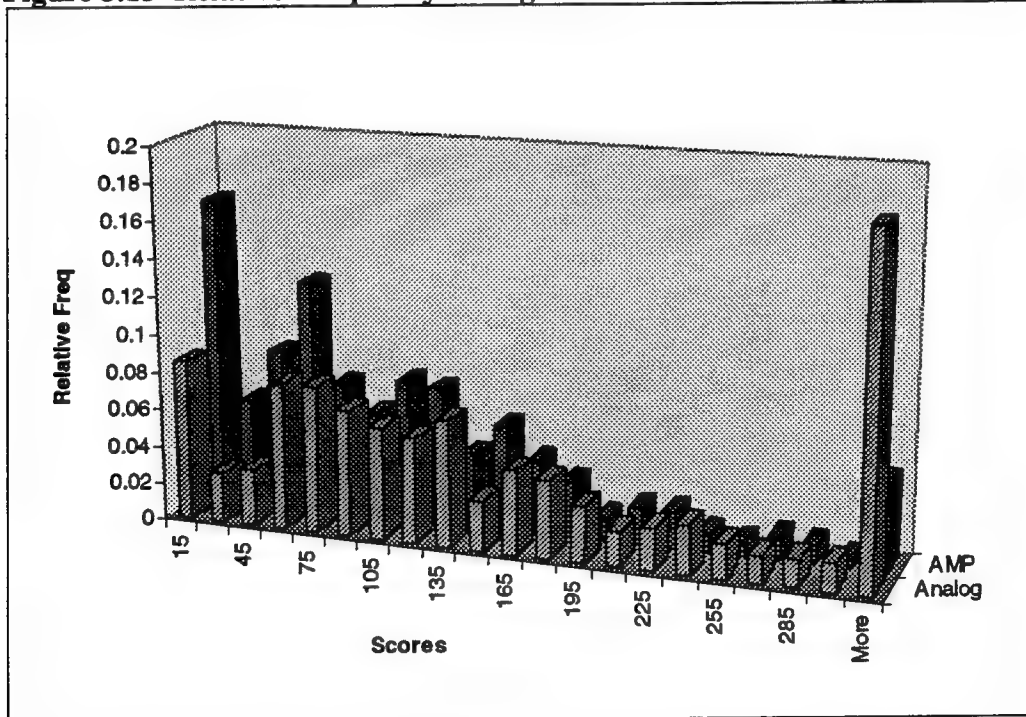**Figure 3.12 Q-Q Plot: AMP vs. Analog (RLD Scores)**

Figure 3.13 Relative Frequency Histograms: AMP vs. Analog (RLD Scores)

From the above results, it is clear that all aircraft cannot be grouped together for RLD scores. Separate control limits and control charts must be applied to each type. A question that remains: Can all aircraft of the same type be charted together? An ANOVA of scores grouped by tail number for each aircraft type can answer this question. Figure 3.14 is the ANOVA table for the analog jets. With a significance level of 0.280, analog aircraft can be grouped together.

Figure 3.14 ANOVA: Analog RLD Scores (by jet)

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	504818.7	8	63102.34	1.235505	0.28049	1.988461
Within Groups	9499788	186	51074.13			
Total	10004607	194				

For the AMP jets, the grouping is not justified for all aircraft. Figure 3.15 shows that not all AMP jet scores come from distributions with the same mean, at a significance level of 0.028. Note that tail number 27 has the largest mean and variance. Figure 3.16 is the ANOVA of the AMP jets with tail number 27 removed. This test shows that all other AMP jets can be grouped together at a significance level of 0.504. Aircraft 27 appears to be an outlier. The possible causes for this condition are numerous. Perhaps the jet had a slowly deteriorating bombing computer, or an intermittent GPS receiver. Maybe the bomb racks on the aircraft had poor springs or sticky doors. The purpose of the AMP modification was to improve bombing and navigation, but tail number 27 shows no obvious advantage over analog jets in bombing. Whatever the cause, because such a drastic difference exists in its mean score, an assignable cause is assumed to exist. Without specific knowledge of the cause, such an assumption is indeed speculative, but the aircraft should be excluded from control limit calculations. Scores obtained in it will still be plotted with other aircraft, but conclusions about the state of control of an aircrew's bombing process should be avoided if the aircraft is involved.

3-15

Figure 3.15 ANOVA: AMP RLD Scores (by jet)

Groups	Count	Sum	Average	Variance
22	36	3897	108.25	37032.31
27	51	9710	190.3922	91189.76
32	62	6073	97.95161	12064.77
40	45	4170	92.66667	7720.864
44	57	4825	84.64912	4174.16
47	9	1000	111.1111	6217.361
48	14	2283	163.0714	23801.61
50	39	3967	101.7179	8313.734
54	11	1130	102.7273	4694.618
63	38	4001	105.2895	7224.968
77	18	1949	108.2778	5601.389
83	17	1034	60.82353	2591.904

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	473235.5	11	43021.41	1.997717	0.0275	1.81355
Within Groups	8291087	385	21535.29			
Total	8764322	396				

Figure 3.16 ANOVA: AMP RLD Scores (by jet, without 27)

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	103738.4	10	10373.84	0.931299	0.504354	1.859011
Within Groups	3731598	335	11139.1			
Total	3835337	345				

Between Aircrew Variability. Similar to the above discussion, aircrew differences may impact control limit computations. It would be advantageous to be able to apply the same control limits to all aircrews but, if differences exist between aircrews, it would be inadvisable to do so. Control limits calculated from aircrews whose scores come from different distributions are usually inflated and less likely to highlight out-of-control situations. Practically speaking, it would be expected that different groups of aircrews exist within which common bombing performance is reflected. Grouping might be based on experience or flying frequency. Within the groups, the scores would be homogeneous enough to support common control limits. Unfortunately, without knowledge of the flying experience and frequency for the aircrews included in the data of this study, tests for the existence of smaller subgroups is impossible.

VLB Results. An ANOVA of the 31 pilots who produced the VLB scores in the data (Figure 3.17) shows no significant difference in the means of scores by different aircrews. All of the pilots can be grouped together when calculating control limits.

With the data grouped by aircrew, it is convenient at this point to address the assumption of a constant bombing process. The control charts of Chapter Four will be most useful in determining if the bombing process is constant; however, a simple plot of average scores vs. number of bombs dropped is also helpful. For VLB scores, this plot (Figure 3.18) shows no trend. Without knowledge of experience levels, care should be taken not to generalize the conclusions from this plot. It merely indicates no short term experience effect (in six months of bombing) on bomb scores. More complete conclusions about the process will be made in Chapter Four.

Figure 3.17 ANOVA: VLB Scores (by aircrew)

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	431374.1	30	14379.14	0.815016	0.747716	1.479904
Within Groups	9738808	552	17642.77			
Total	10170182	582				

Figure 3.18 Scatter Plot: Average Score vs. Bombs Dropped (VLB Scores)
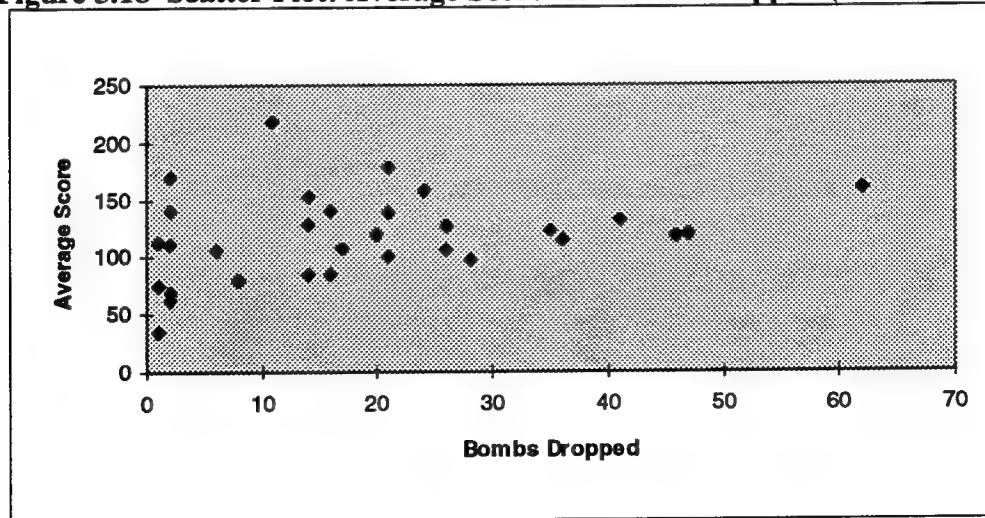
RLD Results. An ANOVA of the RLD scores should be accomplished separately for AMP scores and non-AMP scores since these aircraft were shown to bomb differently. Figure 3.19 illustrates this point. By grouping all scores by aircrew, a significant difference exists in the population, with a significance level of 0.0014. The ANOVAs obtained by splitting the scores into AMP scores and non-AMP scores are depicted in Figure 3.20 and Figure 3.21, respectively. For AMP jets, no significant difference exists between aircrews; however, the non-AMP ANOVA indicates that a difference still exists between aircrews, with a significance level of 0.030. Without more information about the jets or the aircrews, the only conclusion which can be drawn from this result is that bombing performance in non-AMP jets is a function of both aircrew and aircraft. This fact implies that control limits should not be calculated by grouping all aircrews in non-AMP jets together.

Figure 3.19 ANOVA: RLD Scores (by aircrew)

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1584741	30	52824.71	2.008468	0.001396	1.481615
Within Groups	13413510	510	26301			
Total	14998252	540				

Figure 3.20 ANOVA: AMP RLD Scores (by aircrew)

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	360717.1	29	12438.52	1.132786	0.29552	1.503393
Within Groups	3480809	317	10980.47			
Total	3841526	346				

Figure 3.21 ANOVA: Analog RLD Scores (by aircrew)

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2244119	29	77383.4	1.635385	0.029661	1.536765
Within Groups	7760177	164	47318.16			
Total	10004296	193				

In spite of this result, control limits for individual aircrews in specific aircraft quickly become unwieldy and impractical, with 30 aircrews resulting in 630 calculations and 21 control charts for each aircrew with very few observations on each chart. Thus in the next section, control limits are calculated based on all aircrews in non-AMP jets. Chapter Four examines the impact of these control limits on the ability of the corresponding charts to highlight out-of-control conditions.

As a final note, plots of average score vs. number of bombs dropped for AMP and non-AMP RLD data are depicted in Figure 3.22 and Figure 3.23, respectively. Neither shows an indication of short term improvement, and thus (superficially) confirm the assumption of a constant process.

Figure 3.22 Scatter Plot: Average Score vs. Bombs Dropped (AMP RLD Scores)
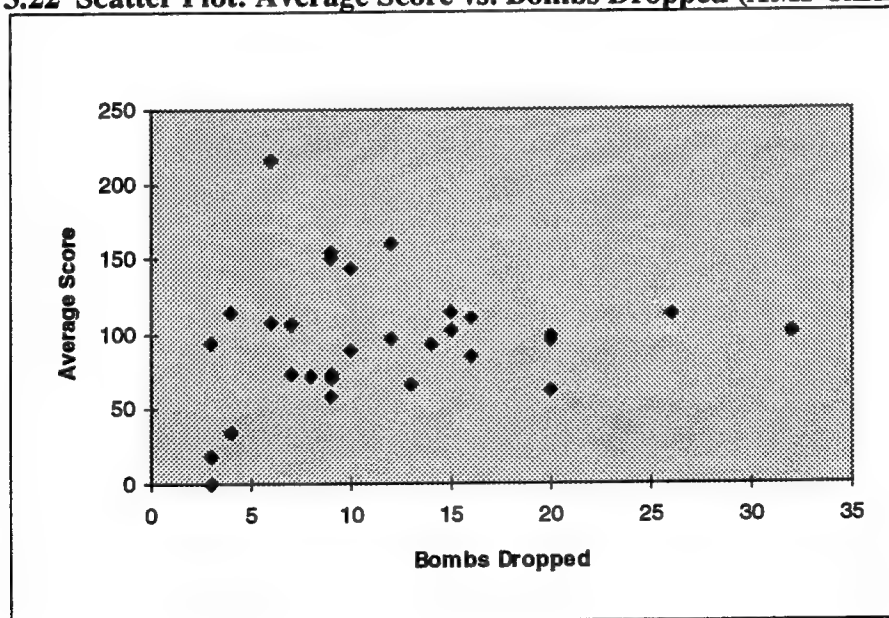
Figure 3.23 Scatter Plot: Average Score vs. Bombs Dropped (Analog RLD Scores)
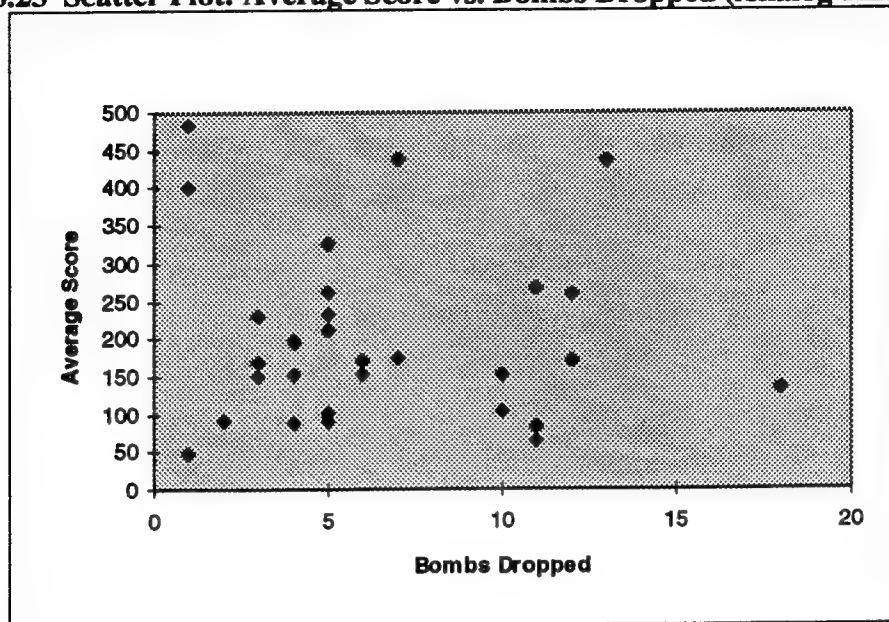
Summary of Preliminary Analysis. It is helpful at this point to summarize significant preliminary results and highlight their implications to the control charting strategy. First, even though the data gave some indication of correlation between first and second passes, XmR charts will be used for all data points. The impact of the correlation on the control limits will be examined in the next chapter. For the sake of completeness, if the correlation severely impacts the control limits, possible ways to handle the situation include: plot first and second passes on separate XmR charts, or plot XmR charts for the residuals of the model of the correlation function, or use a standardized chart based on average scores per sortie.

With regard to how to chart the data, the data gave several indications of non-homogeneity. For VLB scores, ROSY range had a lower mean score than the rest of the population, and should be excluded from control limit calculations, and although it will not be done here, should also be plotted on separate control charts with distinct limits. For RLD scores, AMP jets and non-AMP jets comprise two different groups, thus scores

3-20

for each should be plotted on separate control charts with distinct control limits. In addition, unknown factors caused AMP jet number 27 to have a much higher mean than the rest of the sample; it should be excluded from control limit calculations. For scores grouped by aircrew, in spite of some differences between aircrews in RLD bombing, all aircrews will be used to compute squadron control limits. The next section describes the calculation of control limits for the data following this guidance.

3-2. Calculation of Control Limits

One step remains prior to plotting control charts and applying SPC to bombing proficiency: calculating control limits. This step is the most important to the procedure because control limits are the devices used to monitor the process. Without effective control limits, control charts are incapable of capturing out-of-control situations. This section describes how the preliminary analysis was used to calculate control limits for the data.

As well as excluding outlier data as described in the previous section, out-of-control points in the remaining data **with assignable cause** must also be excluded. This exclusion occurs iteratively; control limits are calculated from all data, then points outside these limits are excluded (or not, based on assignable cause) and control limits are recalculated. This process repeats until no more points plot out of control. The key to the procedure is the determination of the presence of an assignable cause for an out-of-control point. In the Navy study of Chapter Two, the analysts had the opportunity to collect the data in real time. Their presence at the gunnery range allowed the analysts to exclude data points from control limit calculations based on their observations of assignable causes. This method is the preferred procedure, since no guesswork is involved in deciding if an

assignable cause exists. Those squadron weapons officers calculating their own control limits should use this method, if at all possible. That luxury is not present in this study. Decisions must be made whether to exclude out-of-control points based on the author's experience and judgment. For this reason, several different methods of excluding out-of-control points were applied to the data.

First, no points were excluded from the data. Since any assignable cause for an out-of-control point would have to be assumed to exist, this is a justifiable method of calculating control limits. The resulting limits, however, are quite wide, and may be unable to capture many out-of-control situations.

Second, all out-of-control points were excluded. Since any out-of-control point can be assumed to have an assignable cause, this is also a justifiable method. Limits obtained by this method, however, are very tight, and may give too many false alarms to be useful.

Third, only grouped out-of-control points were excluded. Grouped means that the point is immediately preceded or followed (chronologically within one sortie) by another out-of-control point. Since grouped out-of-control points can be assumed to have an assignable cause that is not correctable by the aircrew, this is also a justifiable method.

Fourth, only single out-of-control points were excluded. Single means that the data point is not immediately preceded or followed by another out-of-control point. This method has no real justification, and is included for comparison purposes.

After excluding ROSY scores from VLB data, jet number 27 scores from RLD data, and all scores obtained from non-squadron jets, then applying all four methods described above, the control limits were calculated. These are shown in Figures 3.24, 3.25, and 3.26 for VLB, AMP RLD, and Analog RLD scores, respectively. The most effective limits are those obtained by the third method described above (excluding grouped points). The other three methods produced control limits that were either unjustified in

their calculation, were too wide to capture out-of-control situations, or were so tight as to produce too many indications of out-of-control situations (false alarms). The limits obtained by excluding groups appear to have the best mix of method justification and width. First, points occur in groups due to some factor in the bombing solution not accounted for by the aircrew. It may be unreported winds, the wrong release range, a bad altitude calibration of the system, or many other factors. Single out-of-control points may also be caused by these factors, but may also be caused by the natural variation in an aircrew's bombing performance. Second, the control limits resulting from excluding groups are of a more appropriate width for the variation seen in the data. Any choice of a method is a tradeoff between all of the control limits obtained, and those of the group exclusion method appear to fit the data better than any other method. Chapter Four uses these control limits to plot control charts for several aircrews and examine their indications.

Figure 3.24 VLB Control Limits

Excluding	Xbar	mRbar	LCL(X)	CL(X)	UCL(X)	CL(mR)	UCL(mR)
none	127.2607	120.3351	0	127.2607	447.3522	120.3351	393.2553
all	99.43762	83.33267	0	99.43762	321.1025	83.33267	272.3312
groups	122.4939	116.3755	0	122.4939	432.0527	116.3755	380.315
singles	111.6179	95.96219	0	111.6179	366.8773	95.96219	313.6044

Figure 3.25 AMP RLD Control Limits

Excluding	Xbar	mRbar	LCL(X)	CL(X)	UCL(X)	CL(mR)	UCL(mR)
none	99.21676	91.09494	0	99.21676	341.5293	91.09494	297.6983
all	75.40127	57.87324	0	75.40127	229.3441	57.87324	189.1297
groups	93.68235	84.03871	0	93.68235	317.2253	84.03871	274.6385
singles	88.89552	75.89836	0	88.89552	290.7852	75.89836	248.0358

Figure 3.26 Analog RLD Control Limits

Excluding	Xbar	mRbar	LCL(X)	CL(X)	UCL(X)	CL(mR)	UCL(mR)
none	195.9227	171.1646	0	195.9227	651.2206	171.1646	559.366
all	149.6298	117.6842	0	149.6298	462.6698	117.6842	384.592
groups	177.6085	152.6667	0	177.6085	583.7018	152.6667	498.9147
singles	176.7316	149.2625	0	176.7316	573.7698	149.2625	487.7899

Due to the complex nature of the procedure, it is recommended that a squadron wishing to apply SPC methodology to their bombing process request their command headquarters analysis team to calculate historical control limits for their aircraft type and update them as aircraft modifications occur. The analysis team will have the expertise required, as well as access to historical data needed to compute aircraft control limits. Their control limits will be stable, and are likely to be more accurate than control limits calculated by the squadron. If a squadron finds it necessary to compute their own control limits, appendix A is a guide on how to do so.

4. Control Charts

With data and control limits in hand, all that remains is to compute values for several key control chart properties, then plot the data and examine the results. This chapter begins with a look at control chart properties such as the probability of type I and type II errors, in-control run length, and out-of-control run length. The EXCEL add-in software package BestFit will be used to select a theoretical distribution for the data. The chi-square goodness-of-fit test will be used to confirm the validity of the selection. Properties of this theoretical distribution will then be compared to the results derived empirically. From these calculations and comparisons will come an indication of how effective the control charts are with the data.

Following this discussion comes the final topic, the actual control charts. Control charts will be plotted for several individuals and discussed. Do the charts indicate out-of-control situations? Are they false alarms? Do the charts miss out-of-control situations? What should a squadron weaponeer do when he or she is presented with the information the control charts provide? These and other questions will be addressed in this section.

4-1. Empirical and Theoretical Distribution Comparison

Control charts are the tools used in SPC to monitor and improve a process. The effectiveness of these tools is important information that will determine the success of the SPC effort. Several metrics exist by which control chart effectiveness can be gauged.

This section discusses the four most important metrics to control charts: the probability of type I error, the probability of type II error, in-control run length, and out-of-control run length. In this section, each of these will be defined and then applied to the data to assess the overall effectiveness of the control charts for the data in the study.

Following this discussion is the calculation of the theoretical values of each of the metrics based on the distribution chosen for the data by BestFit and confirmed by the chi-square goodness-of-fit test. A comparison of the theoretical and empirical values will disclose how well the data fit the theoretical distributions, as well as the effectiveness of the control charts.

Control Chart Metrics. As in any statistical application, the indication of the existence of a condition which does not exist in the data is called a type I error. In control charts, type I errors occur when the data indicate an out-of-control situation when the process is actually in-control. The probability of a type I error is called the false alarm rate, and depends on the underlying distribution of the data. Associated with the probability of a type I error is the in-control run length. It is a reciprocal measurement to the probability of type I error which indicates how many samples it will take the chart to produce a type I error.

Conversely, the existence of a condition without indication by the data is called a type II error. In control charts, type II errors occur when the control charts do not capture an existing out-of-control condition. Probabilities of type II errors depend both on the underlying distribution and the type of out-of-control condition that exists. Control charts are much more likely to detect a large shift in a process mean than a small or temporary deviation. For a shift of 1-2 standard deviations in the mean it is not uncommon for a control chart to have a type II error probability in excess of 0.6. This merely indicates that the chart has a low probability of detecting the shift in the first few samples after it occurs. The out-of-control run length is a measure of how many samples

4-2

will be taken before the shift in the mean is detected by the chart. It is a reciprocal measurement to the complement of the probability of type II error, and is regarded as much more important than type II error probability when measuring the effectiveness of a control chart.

Table 4.1, columns 2-4, lists the values of the four metrics empirically derived from the data. For type II errors, an assumed shift of 1.5 standard deviations from the mean is used for comparison purposes. The values for type II error metrics were determined by shifting all scores up by 1.5 standard deviations and measuring the resulting proportion of in-control points.

Table 4.1 Empirical and Theoretical Measures of Control Chart Properties

1	2	3	4	5	6	7
	Empirical			Theoretical		
data set	VLB	AMP RLD	non-AMP RLD	VLB	AMP RLD	non-AMP RLD
mean/standard deviation	122/124	94/95	178/188	122/122	94/94	178/178
upper control limit	432	317	584	432	317	584
P(type I error)	0.0191	0.0176	0.0212	0.0290	0.034	0.038
Run Length (in-control)	52.4	56.8	47.2	34.5	29.1	26.6
P(type II error)	0.893	0.874	0.841	0.905	0.894	0.888
Run Length (out-of-control)	9.3	7.9	6.3	10.6	9.5	8.9

Theoretical Distribution Fit. With empirical information on each metric available, it is useful to have theoretical values with which to compare them. These values will provide general guides for the expected effectiveness of control charts produced from different, but similar data.

BestFit was used to find a preliminary theoretical fit for the data in the study. Each fit provided by the software was rank ordered according to three statistical measures internal to the program: the chi-square statistic, the Kolmogorov-Smirnoff statistic, and the Anderson-Darling statistic. Since no distribution was superior by all three measures, a compromise was required. For all three sets of data (VLB, AMP RLD, non-AMP RLD),

the exponential distribution had the best combination of all three statistical measures, and was thus chosen as the underlying theoretical distribution for the data. Figures 4.1, 4.2, and 4.3 show the fit of the theoretical distribution to each set of data. For each plot, the data is in histogram form, and the exponential distribution (with the mean in parentheses) is a smooth line.
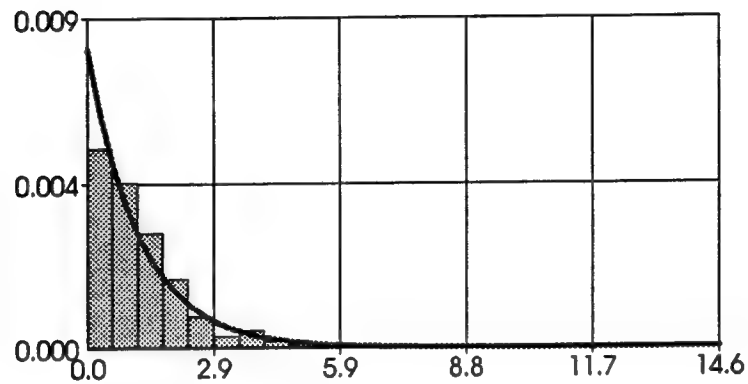
Figure 4.1 VLB Scores vs. Exp(122)

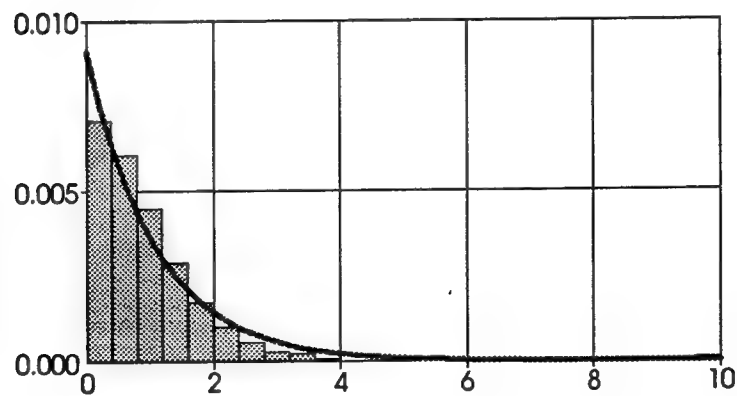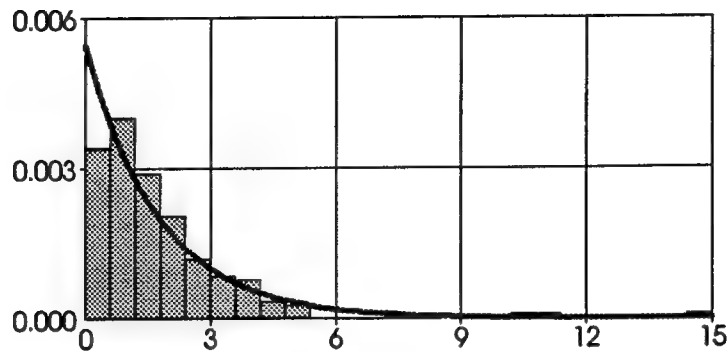**Figure 4.2 AMP RLD Scores vs. Exp(94)**

Figure 4.3 Analog RLD Scores vs. Exp(178)

To confirm the statistical validity of these theoretical distributions, the chi-square goodness-of-fit test was run on each distribution and data set. (The chi-square *value* obtained in BestFit is not the same as the chi-square *statistic* obtained using the chi-square goodness-of-fit test.) Table 4.2 shows the results of these tests. As the table indicates, the test cannot reject the theoretical distributions at the depicted significance level.

Note: While more stringent goodness-of-fit procedures could be applied to determine the absolute best choice of theoretical distributions, such effort is tangential to the present study. It has been shown that control charts are effective and predictable over a wide range of distributions (7:65), and the metrics contained here are used only as guides for predicting control chart performance.

Table 4.2 Chi-square Goodness-Of-Fit Results

data set / distribution	chi-square value	critical chi-square value ($p = 0.05$)	significance level
VLB / Exp(122)	22.99415	24.9948	0.08426
AMP RLD / Exp(94)	17.6013	18.307	0.06207
non-AMP RLD / Exp(178)	12.7184	23.6848	.54881

Table 4.1 columns 5-7 show the theoretically calculated values for the four metrics described. Each was obtained using the exponential distribution with the appropriate

mean and standard deviation. From these values, the performance of a control chart based on the exponential distribution can be described. First, for VLB scores, the chart will make a type I error 2.9% of the time. With an average of 25 VLBs dropped per half, this probability of a type I error results in less than one false alarm per six months per individual. The in-control run length of 34.5 reflects this error rate.

The AMP RLD control chart numbers are very similar to those for the VLB scores. A 0.034 type I error rate, over an average of 13 bombs per half, results in less than one-half false alarm per six month period. (The overall average number of RLD scores per aircrew is 21. Of these, 40% are non-AMP scores and 60% are AMP scores, giving an average of 8 and 13, respectively.) Slightly higher percentages exist for non-AMP RLD scores. Since an average of 8 bombs were dropped by each aircrew in a six month period, the performance of the charts remains about the same as the VLB charts. A 0.038 type I error rate, over an average of 8 bombs per half, again results in less than one false alarm per six month period. It is clear from these results that the correlation between first and second passes was low enough to have very little impact on the control limits; they are not so tight as to produce an unacceptable number of false alarms.

Somewhat more disturbing information is provided by the type II error probabilities for all three data categories. Each type II error probability is near 0.9. In other words, for 90% of the out-of-control points, the chart will fail to recognize it. This fact translates into an out-of-control run length of between 6 and 10. It will take the charts 6-10 consecutive out-of-control points before the chart will highlight the situation. With only 8-25 bombs dropped in a six month period, it could theoretically take quite a long time for the charts to capture an out-of-control condition. Note that these results are for a shift in the process mean of 1.5 standard deviations. Larger shifts are possible, and would be captured earlier by the charts. Smaller shifts would take even longer to capture.

To counter this situation, *run rules* can be instituted. If a run of consecutive scores occurs, either above the center line, below the center line, or with an upward or downward trend, an out-of-control condition may be present. The number of consecutive scores required to signal an out-of-control condition, as well as the adverse effect the use of run rules will have on the false alarm rate, will depend on the process. Section 4.2 looks at charts with and without out-of-control indications before drawing conclusions about the impact of the long out-of-control run length and the use of run rules.

Comparison of Values and Conclusions. By comparing corresponding columns of Table 4.1, several key observations about the effectiveness of the control charts for the data in this study can be made. All empirical type I error probabilities are better than predicted by the theoretical distribution; thus, the control charts for the data can be expected to have an average of less than one false alarm per individual. Empirical type II error probabilities also compare favorably with theoretical values. The next section shows what this result means to the user of the control charts.

4-2. The Control Charts

With the knowledge of how the control charts are expected to perform with the data, there remains nothing left but to plot some individuals and examine their bombing performance. This section examines X and mR charts for several pilots and WSOs. Each aircrew's charts provide information about their bombing process, and all of the charts taken together give insights into the squadron bombing process as a whole.

A tool which helps trace the consistency of an aircrew's bombing is the prediction limit. Prediction limits (upper/central/lower, or UPL/CPL/LPL, respectively) are identical to the control limits previously described in all ways except for two: prediction limits use

only the bomb scores for a specific individual, and they do not exclude any out-of-control points. Because of these elements, prediction limits can be considered loose, individualized control limits. They will prove to be important discriminators when examining control charts for consistency.

Consistent Bombers. By examining the charts and comparing individual prediction limits to squadron control limits, consistent bombers can be identified. These aircrews are those whose scores are steady and whose prediction limits are tighter than squadron control limits. In the data, several consistent bombers exist. In the VLB scores, pilot #2 stands out as a very consistent bomber. He also out-performs the squadron as a whole. His XmR charts are shown in Figure 4.4. On the X chart, the top line is the UCL for the squadron as computed in Chapter Three. The next line is pilot #2's UPL. His personal upper control limit is nearly 100 feet lower than the squadron limit. His CPL is also lower than the squadron average. The plot of the bombs appears to be random about its mean for the most part. An upward shift might be seen during the period between 27 April and 27 May. In the same period he drops his only outlier bomb, which is outside his UPL, but well inside the UCL. The mR chart shows no trend during this or any other time.

For this individual, the job of the weaponeer is simple. First, interview the pilot to find out who he flew with during the questionable period, as well as what aircraft he flew and the ranges he visited. The goal is to find an assignable cause for the diminishing performance. Even if no cause is found, this individual obviously bombs at a better level than the squadron. His techniques could be insightful and worthwhile for everyone to hear. In other words, this pilot would be suitable to teach a classroom session on visual level bombing. If assignable causes were found for the 27 April to 27 May period, they could be excellent tools for distinguishing good techniques from poor techniques, or

highlighting deficient crewmates, jets, and ranges. The squadron stands to learn much from consistent performers like pilot #2.

Inconsistent Bombers. As a contrast to pilot #2, Figure 4.5 shows the XmR charts for pilot #71. The period covered is shorter because pilot #71 dropped three times as many bombs as pilot #2 in six months. In spite of all of this practice, pilot #71's bombing displays classic signs of being out-of-control. On both charts his UPL is above the UCL; his CPL is above the squadron average; five bombs plot out-of-control; four moving ranges plot out-of-control. While not shown, the remainder of his scores for the six month period are very similar to those shown.

For this individual, the job of the weaponeer is more difficult. Pilot #71's problem is simply stated: He is not getting to the proper release point consistently; however, the weaponeer must investigate all possible causes for the problem. It could be the squadron's crewing policy; he needs more consistent crewing to develop proper habit patterns on the range. Maybe he's ignoring something he should be monitoring on the range; his previous instructor always told him when he was slow; now, without that instructor, he isn't watching his airspeed. The possible causes are many, and the weaponeer knows most of them. The key is to replace the poor habit patterns with good habit patterns, so he can get to the proper release point more consistently. Pilot #2 could probably teach pilot #71 some valuable lessons in a classroom, and he should not be spared the responsibility of doing so. Squadron's flourish when experience is shared.

Control Chart Effectiveness. Previously, it was predicted that the control charts would take 6-10 out-of-control points before they could highlight the situation. This problem wasn't apparent in the above charts, but it does occur in the data. Figure 4.6 shows the XmR charts for WSO #69 in analog jets. Several items are noteworthy on these charts. First, the UCL is very high. In Chapter Three, it was stated that even though differences exist between analog jets, they would be plotted together to see the
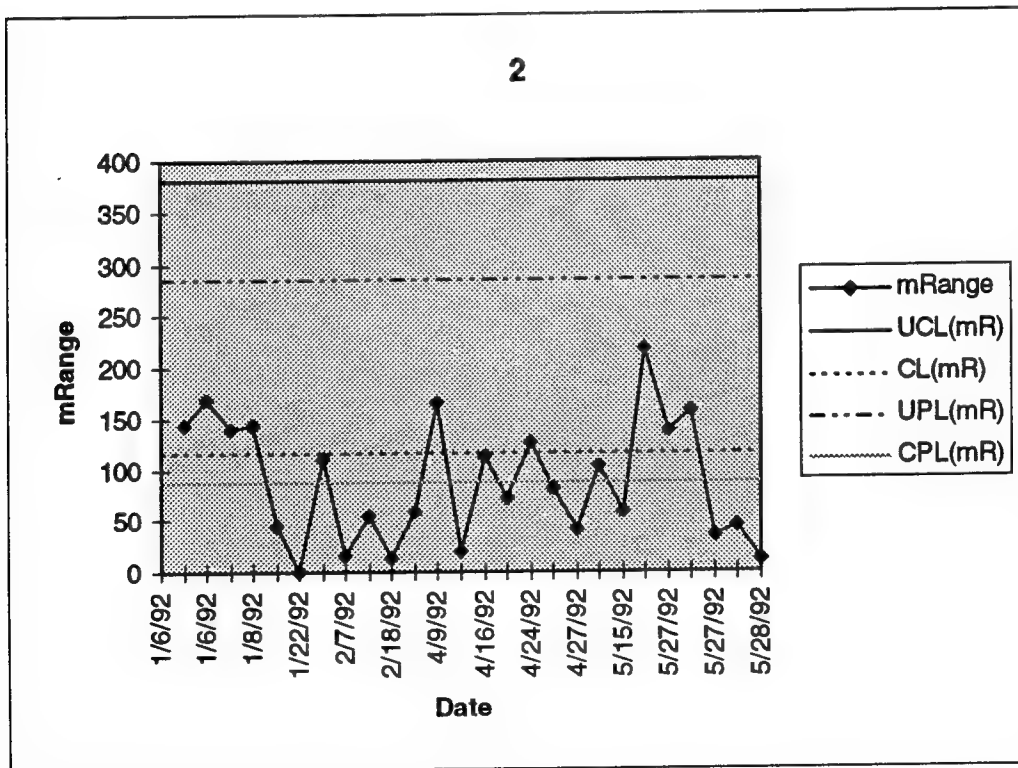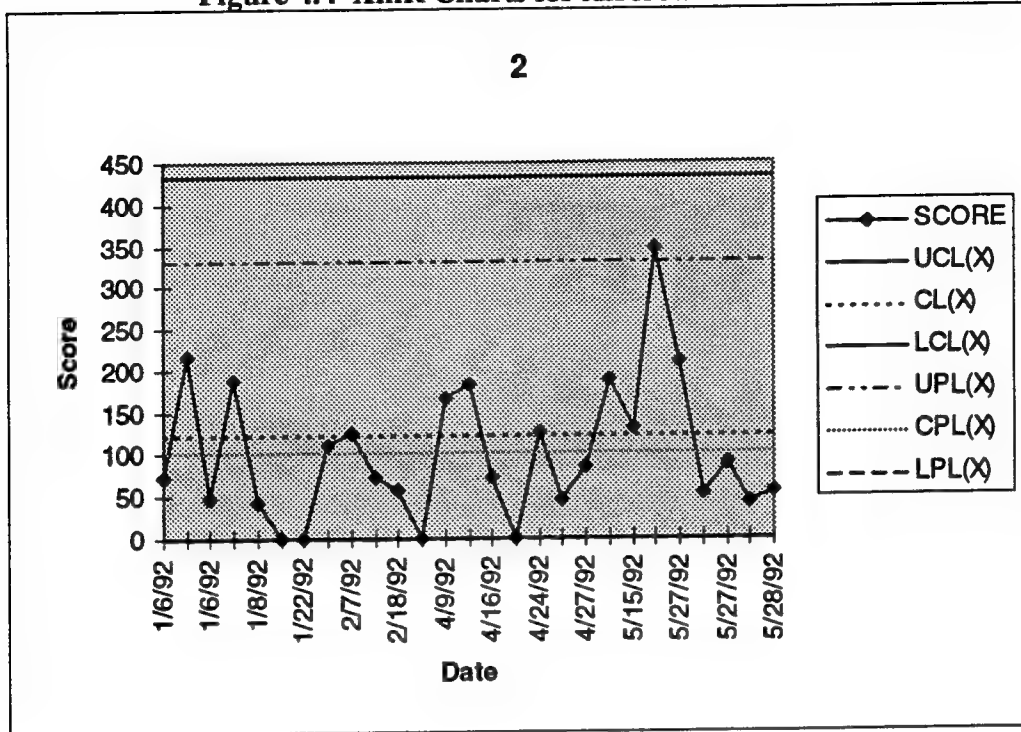
Figure 4.4 XmR Charts for Aircrew Number 2

Figure 4.5 XmR Charts for Aircrew Number 71
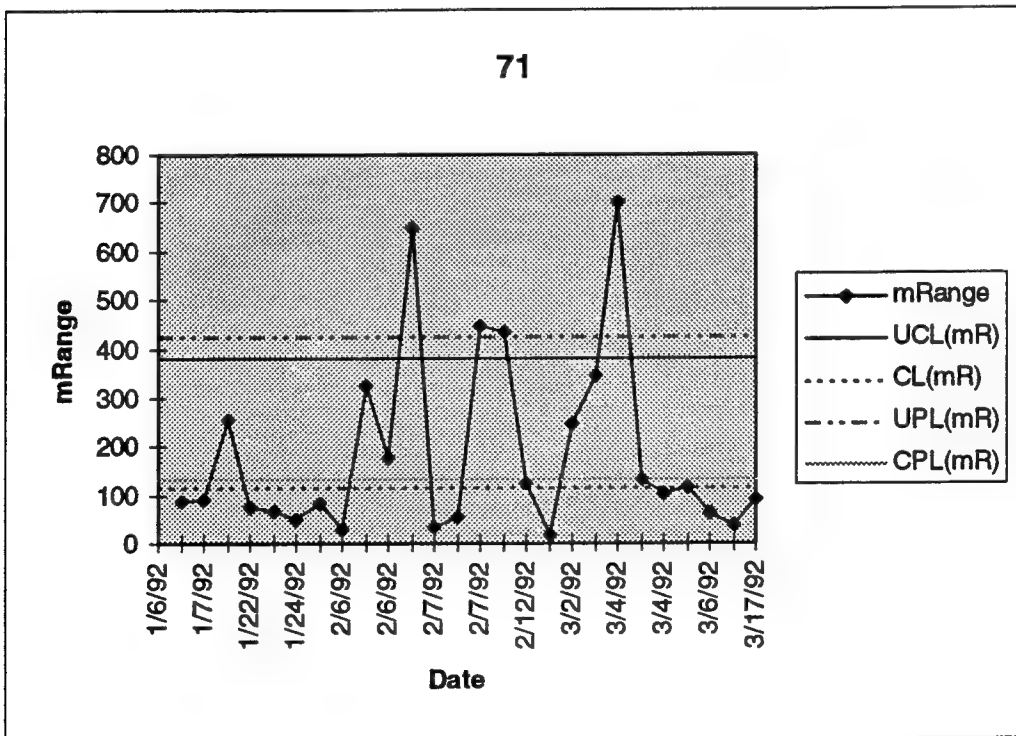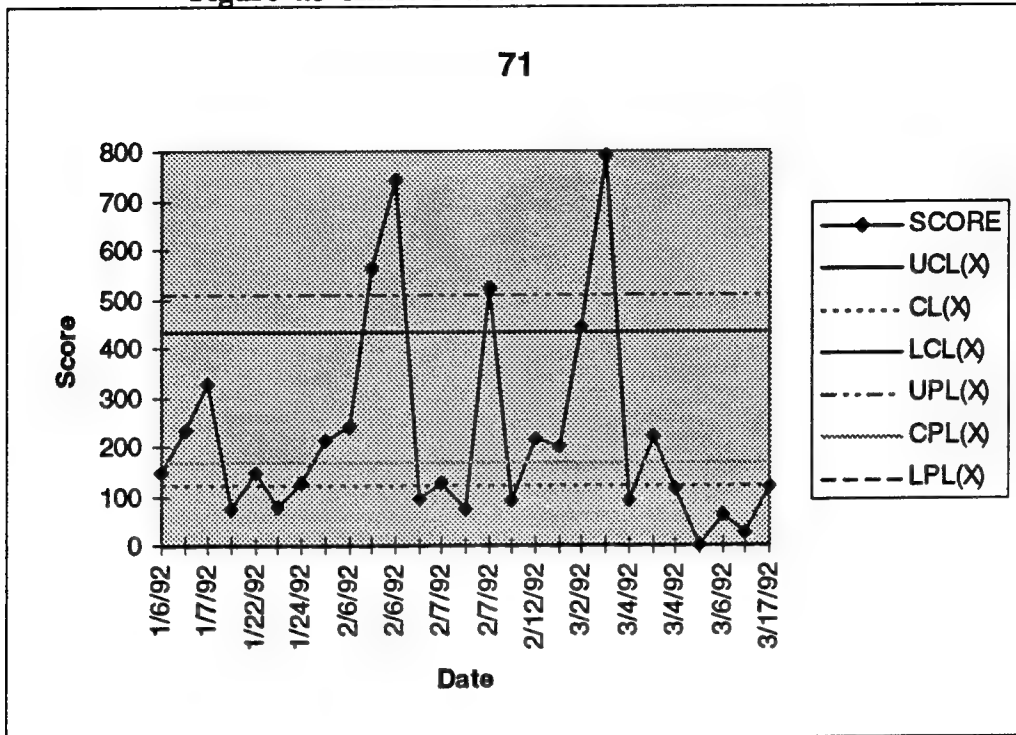

71

Score chart legend: SCORE, UCL(X), CL(X), LCL(X), UPL(X), CPL(X), LPL(X)

71

mRange chart legend: mRange, UCL(mR), CL(mR), UPL(mR), CPL(mR)
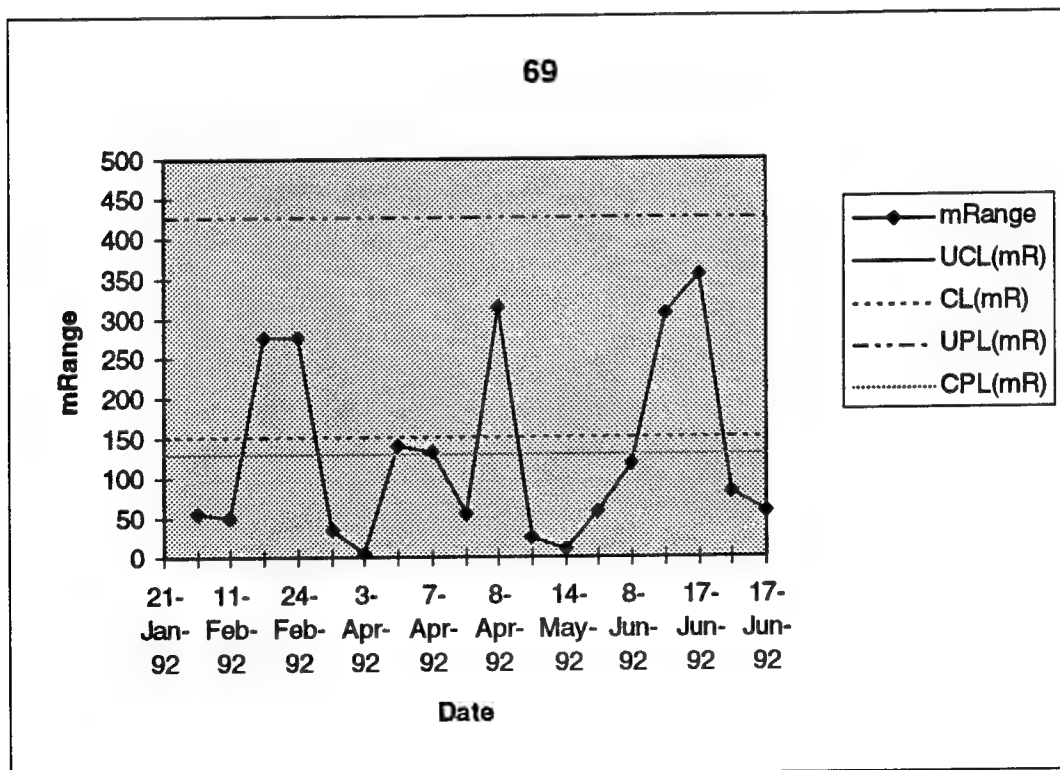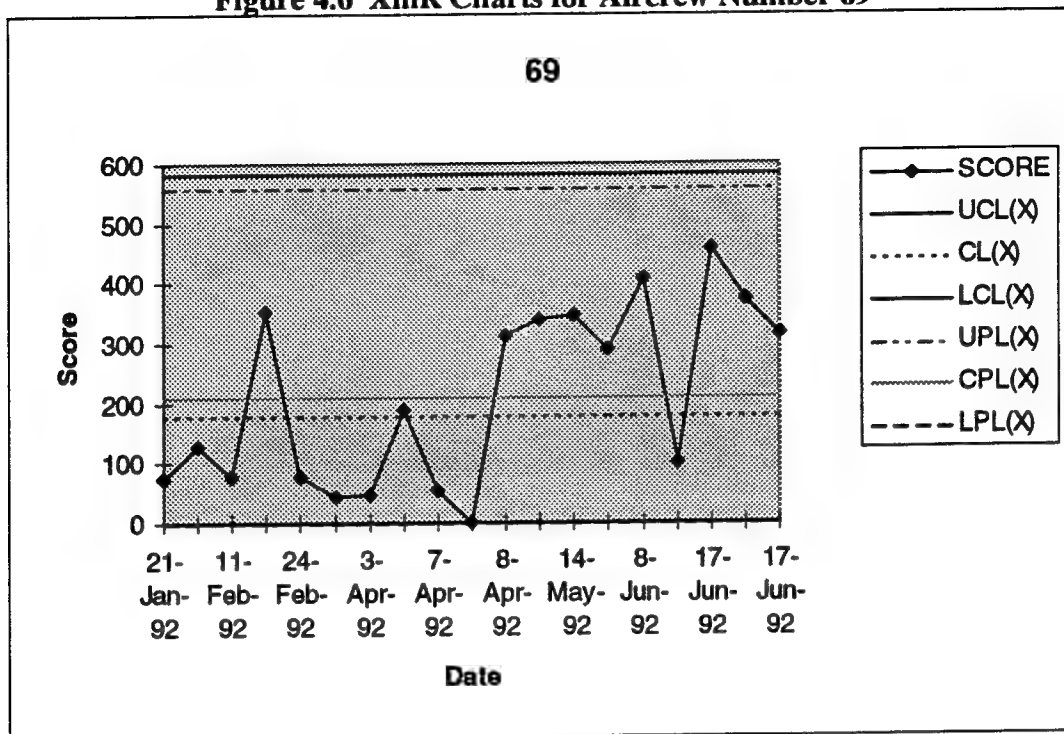
4-11

Figure 4.6 XmR Charts for Aircrew Number 69

69

69

4-12

result. The wide control limits are the result of grouping the analog jets together. Second, the long out-of-control run length (6.3), coupled with the low average number of bombs per individual per six month period (8), means that a good chance exists that the chart will not capture out-of-control situations within a six month period. Run rules can be used to counter these conditions. For example, an examination of WSO #69's scores shows that a shift may have occurred around 8 April. Before this time, scores were consistently below the squadron average. After this time, all but one score were well above the mean. This run of 8-out-of-9 can be interpreted as an out-of-control situation, even though no points actually plot out-of-control. Treatment of this situation is identical to any other out-of-control situation. This example serves to illustrate how control charts can be effective in spite of long out-of-control run lengths and/or wide control limits.

Summary of Control Charts. Even though only a few examples are included in the above discussion, their indications appear repeatedly in the data: as would hopefully be expected, out-of-control situations occur less frequently than do situations of consistent, in-control bombing performance; virtually every control chart confirmed the assumption of a constant process, with random fluctuations about a mean; the low level of correlation present in the data caused no appreciable tightening of the control limits; prediction limits for each aircrew indicated whether he bombed better than average, average, or worse than average. This discrimination is the key to improving the bombing process of the squadron as a whole. By highlighting those who need help, as well as those who can help them, the squadron can develop a program of instruction that should eventually raise the standard for the bombing of all aircrews. It is clear that control charts are an effective tool for obtaining this improvement.

5. Conclusions and Recommendations

The original goal of this research was to determine if SPC techniques could provide a useful tool for managing ABP. SPC methods are designed to monitor a process by measuring the variability in that process. By considering the act of dropping bombs on a range a process with measurable output, the well established methods of Statistical Process Control can be applied. Improvements to the process are made possible as causes of variation in the process are identified and removed.

The question remains: Do control charts provide squadron supervisors with better means to monitor the bombing performance than the qualification program in existence? To answer yes, the capabilities of control charts must go above and beyond the current qualification program. To be beneficial, a new tool should aid in the identification of problems **and** potential problems in bombing within the squadron. In the past, models have been developed for the relationship between experience, or flying frequency, and bombing proficiency. These models have not provided weaponeers with near real-time solution methods to capture potential problems or to improve bombing in the squadron. SPC provides those solution methods. Preliminary research in the area of Naval gunfire support has shown that SPC methods can be useful in reducing training costs, as well as identifying gun crews and ship turrets which are not performing to expectations. Parallels between NGFS and ABP abound; thus SPC applications to ABP should be similarly productive.

In Chapter Three, the extremely important issue of how to chart data from the bombing process was discussed. While the results of the statistical tests used to decide how to plot each aircrew, aircraft, and range were specific to the data, generalizations are

plausible. Since the goal was to track individual aircrews, it seemed obvious to chart each aircrew separately. First and second passes should not differ much, and aircraft and range differences should not matter, so they should all be plotted together on an individual's chart. The data supported these hypotheses for the most part. Chapter Four evaluated the impact of the difference between the jets, and noted how other properties of the control chart can serve to overcome the problems caused. The aggregated results indicate that in spite of some degree of heterogeneity in the data, charts based on an assumption of homogeneity are still effective as aids in managing the process.

The most important thing discovered is that control charts have something to say about the bombing process. Squadron supervisors can use them to track bombing within the squadron, and highlight the consistent bombers as well as the erratic ones. Being able to distinguish the two is valuable information for many squadron decisions: who to upgrade, who to send to a high-visibility bombing competition, and who to sit in a classroom to learn good habit patterns. While simply plotting the data as a time series can provide almost as much of the same information, the addition of control limits and prediction limits to those plots adds considerable capabilities. Not only do they give indications of when an individual's performance is outside of what is expected from the bombing process, but they also provide predictions of the future performance of individuals (prediction limits) and the squadron as a whole (control limits). These capabilities are not provided by the current monitoring system; the added value makes this research worthwhile.

It is the recommended that this thesis and the accompanying EXCEL macro (see Appendix B for instructions in its use) be forwarded to Air Combat Command for evaluation. The author will be flying the F-117 for three years upon graduation from AFIT, and will be in an excellent position to make improvements to and/or trial implementations of the methodology.

APPENDIX A - How To Compute Control Limits

This appendix is a guide to computing control limits from existing data using EXCEL. It is recommended that historical control limits be calculated by trained analysts at ACC headquarters for each aircraft type. If this is not possible, the following discussion will guide the initiated through the calculation of control limits for use on X and mR charts. For those persons with a statistics background, more advanced control limit information can be obtained from [reference 6] or [reference 7].

1. Gather data:
   - Collect enough data to make the calculations representative. Scores should come from all experience levels, all bombing ranges normally used, all squadron owned aircraft, and as long a period as possible.
   - The information required for each point is: aircrew name, date, aircraft, range, and score.
   - Quite a few points have the potential of being excluded, so collect enough scores so that the remaining points are still representative.

2. Set up a spreadsheet:
   - The first row should be labels. Beginning in column A: name, date, aircraft, range, score, mRange, Out-of-Control?, Xbar, mRbar, LCL(X), CL(X), UCL(X), CL(mR), UCL(mR).
   - The second row, columns H-N should contain formulas corresponding to the labels above them:
     - column H [Xbar] "=AVERAGE(E2:Ex)" the average of the scores, where x is the row number of the last score in the data set

- column I [mRbar] "=AVERAGE(F2:Fx)" the average of the moving ranges
- column J [LCL(X)] "=MAX(0,$H$2-2.66*$I$2)" the X chart lower control limit for sample size of 2
- column K [CL(X)] "=$H$2" the X chart center line is the average of the scores
- column L [UCL(X)] "=$H$2+2.66*$I$2" the X chart upper control limit for sample size of 2
- column M [CL(mR)] "=$I$2" the mR chart center line is the average of the moving ranges
- column N [UCL(mR)] "=3.268*$I$2" the mR chart upper control limit for sample size of 2

- Beginning in row 2, the first five columns of the spreadsheet should contain the data - aircrew name, date, aircraft, range, and score. Sort by name, then chronologically, then by range.
- The sixth column should contain the moving range between successive scores. Use the formula: "=ABS(E4-E3)", where the first number (4 here) is the current row, and the second number (3 here) is the row above. The first score for each aircrew should have no moving range (leave blank).
- The seventh column should contain flags indicating the out-of-control condition for the specific row (score and/or moving range). Use the formula: "=IF(E4>$L$2,IF(F4>$N$2,"both","X"),IF(F4>$N$2,"mR"," "))", where all column E and column F references are to the current row. Drag this formula down the entire length of the data set. The flag will show "X" if only the score is out-of-control, "mR" if only the moving range is out-of-control, and "both" if both score and moving range are out-of-control.

A-2

3. Calculate trial control limits: the spreadsheet does this automatically. They are shown in row 2, columns J-N, under the appropriate heading.

4. Exclude points:
   - Decide on the type of out-of-control points to exclude. If you were involved in the collection of the data, or have information on assignable causes for all out-of-control points, exclude those points with assignable cause. If no information is available, several exclusion plans are available:
     - all - produces the tightest control limits.
     - singles - produces tight control limits.
     - groups - produces comfortable control limits, recommended method.
     - none - produces the widest control limits, may be too loose.
     - other - mix and match or come up with your own plan. Make certain it has some foundation in the exclusion of points with assignable cause.
   - Look down the list of scores, and delete the entire row (shifting up) of the points to be excluded from the calculations.
   - Do not delete all of row 2, as it contains formulas beyond the first 7 columns. (Delete only the columns A-G, shifting up).
   - Drag the moving range formula into the cell showing "#REF!" in the mRange column from above or below, if required. This should clear up all other "#REF!" flags.
   - If excluding an aircrew's first bomb, clear the new first bomb moving range cell.

5. Recalculate control limits: again, the spreadsheet does this automatically.

6. Iterate from 4 until no more points qualify for exclusion. The resulting control limits are in row 2, columns J-N.

A-3

7. Apply control limits to data in your copy of SCORES.XLS (see Appendix B). Assess suitability of exclusion method by examining the control charts for the number of out-of-control points plotted. If an excessive number are shown, the control limits may be too tight. If very few are shown, the control limits may be too loose. Either case may require a new method of excluding points from the calculations above.

8. See sample spreadsheet (Figure A.1) for layout.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Crew #	Date	Range	Tail	SCORE	mRange	Control?	Xbar	mRbar	LCL(X)	CL(X)	UCL(X)	CL(mR)	UCL(mR)
2	7	2/11/92	Tain	071	62	328		177.6085	152.6667	0	177.6085	583.7018	152.6667	498.9147
3	7	2/14/92	Cowde	120	390	260								
4	7	2/14/92	Wainf	120	130	224								
5	7	3/26/92	Jurby	071	354	70								
6	7	4/3/92	Roseh	016	284	115								
7	7	4/13/92	Tain	010	169	169								
8	7	4/13/92	Tain	010	0	180								
9	7	4/27/92	Vleh	016	180	54								
10	7	5/11/92	Jurby	015	126	34								
11	7	6/2/92	Wainf	120	160	915								
12	7	6/2/92	Wainf	120	1075		both							
13	8	2/3/92	Wainf	039	56	144								
14	8	2/3/92	Wainf	039	200	133								
15	8	2/11/92	Roseh	120	67	2								
16	8	2/12/92	Tain	120	69	12								
17	8	2/14/92	Tain	016	57	15								
18	8	3/18/92	Wainf	046	72	10								
19	8	3/18/92	Wainf	046	82	156								
20	8	3/18/92	Wainf	046	238	238								
21	8	4/10/92	Tain	010	0	0								
22	8	5/12/92	Tain	120	0	0								
23	8	5/12/92	Tain	120	64	64								
24	9	1/8/92	Roseh	046	89	64								
25	9	1/8/92	Tain	046	153	153								
26	9	1/8/92	Tain	046	0	0								
27	9	1/22/92	Holb	046	0	0								
28	9	1/22/92	Holb	046	0	154								
29	9	1/22/92	Holb	046	154	88								
30	9	1/22/92	Holb	046	66	79								
31	9	2/20/92	Roseh	119	145	122								
32	9	2/20/92	Roseh	119	23	69								
33	9	3/18/92	Wainf	119	92	69								
34	9	3/18/92	Wainf	119	77	15								
35	9	3/18/92	Wainf	119	154	77								
36	9	3/18/92	Wainf	119	108	46								
37	9	3/18/92	Wainf	119	176	68								

APPENDIX B - How To Run SCORES.XLS

This appendix is a guide for how to use EXCEL 5.0 and the SCORES.XLS workbook to apply SPC and control charts to aircrew bombing.

1. What it is: SCORES.XLS is a workbook containing several macros written for EXCEL 5.0 that automates control charting and data entry/editing/deletion.

2. Requirements:
   - 286/386/486/Pentium running EXCEL 5.0 for Windows
   - enough RAM to run Windows 3.1
   - 200K disk space (for the macro, more for the database)
   - 1.44MB 3.5" disk drive

3. Suggested:
   - 486/33 minimum
   - 8 MB RAM
   - 4 MB free hard disk space

4. Installation: copy the file SCORES.XLS to directory in which you wish to store your bomb score database.

5. Getting Started:
   - in EXCEL, open the file SCORES.XLS. Immediately SAVE AS another file name (the name you wish to apply to that portion of your bomb score database.) Each workbook should contain data for only one bombing event for one type of aircraft for one type of aircrew (i.e., RLD scores for AMP F-111Es for WSOs.)
   - When the program asks if it should re-establish links, click NO.

- The program begins on the Main Screen with an empty database. A menu item entitled Bombing appears in the menu bar. An item under this title is:

  - Return to Main Screen

  This item returns the program to the Interface screen (the one showing now). Also on the Main Screen are three buttons:

  - Select Aircrew

  - Edit Control Limits

  - Delete Aircrew

  The purpose of each button will become clear shortly.

6. Running the macro:

  - Click "Edit Control Limits" and enter the control limits to be applied to the database/control charts (see Appendix A for instructions in computing them).

  - Click "Select Aircrew" to begin entering the first aircrew's information. If the aircrew is not already in the workbook, type the name in the drop down/edit box and follow instructions to add the name to the workbook. The program goes to the aircrew's data page.

  - Click the "Enter Data" button to add scores to the aircrew's record. Enter information when the dialog box appears. Click "Add" to add the information. Click "Done" or "Cancel" after "Add"ing information for the last score. The program adds the information to the data sheet in chronological order. (If you have previous information for the aircrew in a format that EXCEL can interpret and translate, block copy/paste procedures may be used to transfer this information into the spreadsheet. Use standard EXCEL unprotect/protect procedures before and after to avoid corruption of data on the spreadsheet.)

- At this point you can click Bombing/Return to Main Screen to input another aircrew's information, or you can click on the "Plot" button to chart the current aircrew's information. Before clicking "Plot", enter the first and last bomb to plot in the upper right edit boxes. Use the line number of the first and last bomb in the range. After selecting "Plot", you will be on the X chart page. Click Bombing/Go to mR chart to view the mRange chart for this aircrew, or click Bombing/Return to Main Screen to enter more data.

- From the Main Screen, you can click the "Delete Aircrew" button to delete a departing aircrew from the database (or one entered in error).

7. What information do the charts give:

- On each chart are upper and lower control limits (UCL and LCL respectively, lower limit for X chart only). These are the limits for the squadron. If any points plot outside these limits, it should be examined for out-of-control indications (interview the aircrew, find out who was with him, what jet he was in, what ranges he visited.) The center line (CL) indicates the squadron average.

- The plots also show upper and lower prediction limits and a center prediction line (lower limit for X chart only). The prediction limits are based on the information on the specific aircrew's spreadsheet (past performance). As well as providing prediction capability (how he should bomb in the future), the prediction limits serve to highlight the aircrew as bombing better/worse than the rest of the squadron (on whom the control limits are based). If an aircrew's prediction limits are tighter than the control limits, he is bombing better than the squadron as a whole. If they are looser, he is worse. Thus, even with no out-of-control points on the chart, it can still give indications of the aircrew's bombing proficiency relative to the rest of the squadron.

- Look for trends in the plots, like a steadily increasing or decreasing sequence of bombs, or a sudden shift (large moving range) to a higher or lower level. Any trends other than apparent randomness about the CPL can indicate an out-of-control situation and should be investigated (interview the aircrew, find out who was with him, what jet he was in, what ranges he visited; any possible cause for the trend which can be corrected or taught to other members of the squadron.)
- Use consistent bombers with tight prediction limits to teach others good habit patterns on the range.
- Highlight erratic bombers with wide prediction limits for additional training in bombing techniques.

8. Use standard EXCEL printing procedures to obtain hard copies of charts.

9. Use standard EXCEL exit procedures to end the program. Save entered data often while running the program, as well as when exiting.

10. Limitations:
- Each spreadsheet can support over 200 aircrews, but RAM may further limit this size.
- A workbook should contain only one bombing event. The only events supported by the accompanying analysis are VLB/RLD events.

11. Improvements: for the initiated, several pages of macro code are hidden within the program. Use standard EXCEL unhide procedures to view them. Take great care to understand all of the implications before changing something in the program. Keep the original SCORES.XLS unchanged in case copies become corrupted.

Bibliography

1. Bailey, Michael, John Bowden, and Alexander J. Callahan. "Managing Ship Performance of Naval Gunfire Support Using Statistical Process Control," <u>Military Operations Research</u>: 5-11 (Summer 1994).

2. Cedel, Lt Col Thomas E., and Lt Col Ronald P. Fuchs. <u>An Analysis of Factors Affecting Pilot Proficiency: Final Study Report, 1984-1986</u>. Washington: Fighter Division, Directorate of Theater Force Analyses, Air Force Center of Studies and Analyses, The Pentagon, 1986 (AD-B109757L).

3. Chambers, J. M., W. S. Cleveland, B. Kleiner, and P. A. Tukey. <u>Graphical Methods for Data Analysis</u>. Belmont, California: International Group, Boston, Massachusetts: Duxbury Press, 1983.

4. Lewis, P. A. W., and E. J. Orav. <u>Simulation Methodology for Statisticians, Operations Analysts, and Engineers</u> (Vol I). Pacific Grove, California: Wadsworth & Brooks/Cole, 1989.

5. Makridakis, Spyros G., Steven C. Wheelwright, and Victor E. McGee. <u>Forecasting: Methods and Applications</u> (Second Edition). New York: John Wiley & Sons, 1983.

6. Montgomery, Douglas C. <u>Introduction to Statistical Quality Control</u>. (Second Edition). New York: John Wiley & Sons, 1991.

7. Wheeler, Donald J., and David S. Chambers. <u>Understanding Statistical Process Control</u> (Second Edition). Knoxville, Tennessee: SPC Press, Inc, 1992.

VITA

Captain Kirk G. Horton was born and raised in West Milford, NJ where he graduated from West Milford High School in June 1981. He then attended Stevens Institute of Technology in Hoboken, NJ, participated in Air Force ROTC, and earned his Bachelor's of Engineering in Electrical Engineering/Computer Science and his commission in the US Air Force in May 1985. Upon graduation, he attended flight training at Laughlin AFB, TX and was subsequently assigned to the 16th Tactical Reconnaissance Squadron, Shaw AFB, SC from January 1988 through November 1989 to fly the RF-4C. In July 1989, Captain Horton married the former Susan Pringels of Sumter, SC. During his stay at Shaw, Captain Horton was a flight lead, and was qualified in every aspect of RF-4C employment.

In March 1990, Captain Horton was assigned to the 79th Tactical Fighter Squadron, RAF Upper Heyford, UK to fly the F-111E. From January 1991 through March 1991, Captain Horton flew 15 night combat sorties into Iraq, and was awarded the Distinguished Flying Cross and the Air Medal for those missions. During the remainder of his stay at Upper Heyford, Captain Horton became a multi-ship flight lead and instructor pilot. He participated in many NATO exercises and visited virtually every bombing range in western European airspace.

Upon leaving Upper Heyford in May 1993, Captain Horton was given the opportunity to attend the Air Force Institute of Technology, Wright-Patterson AFB, OH to earn his Master's of Science Degree in Operations Research (Operations Analysis) from the Graduate School of Engineering. Upon graduation in March 1995, Captain Horton will be assigned to Holloman AFB, NM to fly the F-117 Stealth Fighter.